C U R S O S E C O N G R E S O S

Santiago de Compostela
2025

30th Congress of the International Society for Forensic Genetics

Edited by
John M. Butler
Ángel Carracedo
María Victoria Lareu
Ana Freire-Aradas
Ana Mosquera-Miguel
Christopher Phillips
María de la Puente

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

30th ISFG Congress

CURSOS E CONGRESOS DA
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
N.º 278

30th Congress of the International Society for Forensic Genetics

Santiago de Compostela, 9-13 September 2024

PROCEEDINGS

EDITED BY
John M. Butler
Ángel Carracedo
María Victoria Lareu
Ana Freire-Aradas
Ana Mosquera-Miguel
Christopher Phillips
María de la Puente

2025
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Edita
Edicións USC
Campus Vida
15782 Santiago de Compostela
usc.gal/publicacions

Maqueta
Tania Sanmartín
Imprenta Universitaria

DOI: https://dx.doi.org/10.15304/cc.2025.1869

Index

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 23-26
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Introduction

The 30th Congress of the International Society for Forensic Genetics (ISFG) took place in Santiago de Compostela, Spain, in the Auditorio de Galicia. Attendees enjoyed 15 pre-congress workshops, 59 oral presentations, and 438 poster presentations. We are grateful to all our presenters who brought such excellent scientific content. ISFG 2024 had 968 registered participants from 63 countries, with the top five countries being the USA (99 participants), China (86), Spain (80), Germany (77) and Italy (60).

The Congress was planned and organized by the ISFG Board and Scientific Program Committee (John M. Butler, Walther Parson, Leonor Gusmão, Marielle Vennemann, Corina Benschop, Sascha Willuweit and Lourdes Prieto) in cooperation with the local organizing committee (Ángel Carracedo, Maviki Lareu, Ana Mosquera, Ana Freire-Aradas, María de la Puente and others from the University of Santiago de Compostela). Support from Cíntia Alves, ISFG Secretariat, and associates from Sebgar, the ISFG 2024 Scientific Secretariat, were crucial for the success of this Congress. All the photography, by Adolfo Enríquez, has allowed memories to be preserved following the gathering.

There were 15 pre-congress workshops (in which some attendees participated in more than one) involving 532 participants from 45 countries. These workshops provided opportunities to learn from some of the very best researchers and practitioners in forensic genetics:

Full-Day Workshops:

- "Forensic Genetic Genealogy" – Daniel Kling and Andreas Tillmar
- "Mixture Interpretation (Basic & Advanced Sessions) " – Peter Gill, Corina Benschop, Helen Johannessen
- "Bioinformatics MPS (Basic & Advanced Sessions) " – Jerry Hoogenboom and Jonathan King
- "Kinship Statistics and Pedigree Analysis (Basic & Advanced Sessions) " – Thore Egeland and Magnus Vigeland

Half-Day Workshops:

- "Body Fluid Identification" – Titia Sijen, Cordula Haas, Margreet van den Berge
- "Mitochondrial DNA" – Walther Parson and Charla Marshall
- "Disaster Victim Identification" – Carlos Vullo and Lourdes Prieto
- "Forensic Database Advisory Board and Ethical Considerations for Forensic Genetic Frequency Databases" – Maria Eugenia D'Amato, Yann Joly, Vanessa Lynch, Nathan Scudder, John M. Butler (filling in for Martin Zieger)
- "Ancestry/Appearance & Phenotyping" – Antonia Heidegger and Catarina Xavier
- "Activity Level Evaluative Reporting" – Bas Kokshoorn, Yvonne van de Wal, Bart Aarts
- "Y-Chromosome Analysis and Statistics" – Lutz Roewer
- "Y-Chromosome Interpretation: Extended/New Methodologies on Y-SNPs and Y-STRs" – Arwin Ralf and Sofie Claerhout

A summary of the feedback received on these workshops can be found on the ISFG website. ISFG members may also access many of the workshop presentations and plenary lecture slides on the Members Area/ Education part of the website.
The Scientific Program Committee, consisting of the seven members of the ISFG Board, along with Congress President Ángel Carracedo and Congress Vice-President Maviki Lareu, reviewed 620 abstracts when they met on April 23-24, 2024.
A decision to limit submitters to a single presentation led to the rejection of a few abstracts where authors had multiple submissions. In the end, 59 oral presentations were selected along with over 500 invited poster presentations. Unfortunately, some of those invited did not register and were therefore removed from the listing of posters. When the ISFG 2024 meeting was held in September 2024, there were 438 posters presented on 14 electronic poster screens. This was the first ISFG meeting utilizing electronic posters, and important lessons were learned from this format in terms of advantages and disadvantages.

The invited plenary speakers at ISFG 2024 included Charla Marshall (USA), the ISFG Scientific Prize winner in 2022, along with Manuel Esteller (Spain), María Martinón (Spain), Duncan Taylor (Australia) and Andreas Tillmar (Sweden). These 45-minute plenary presentations introduced session topics on DNA technology, phenotyping, population genetics, interpretation and statistics and genetic genealogy.

A special panel discussion was held in honor of Peter M. Schneider on the topic of DNA transfer and activity level reporting, which included presentations by Ángel Carracedo (Spain), Mariya Goray (Australia), Bas Kokshoorn (the Netherlands) and Peter Gill (Norway).

ISFG travel grants were awarded to Katherin Barrionuevo, Emily Bibbo, Kayli Carrillo, Shivani Dixit, Franco Marsico, Donna-Lee Martin, Masinda Nguidi, Phuong Pham, Jessica Watson and Olivia Yugovich.

During the closing ceremonies, awards were given for the best oral presentation and the best poster presentation. Miguel Boullon-Cassau, from the University of Santiago de Compostela (Spain), was awarded the best oral presentation on "Exploring Legal Age Estimation Using DNA Methylation". This prize was selected by a committee chaired by Mecki Prinz (USA) with Maja Sidstedt (Sweden), Iva Gomes (Portugal), Mariya Goray (Australia), Jan Fleckhaus (Germany), Fabio Oldoni (USA) and Hwan Young Lee (South Korea). Caitlin McDonald from Flinders University (Australia) received the prize for best poster presentation "Enhancing PCR Efficiency with Machine Learning". This prize was selected by a committee chaired by María de la Puente (Spain) with Athina Vidaki (Netherlands), Ana Freire-Aradas (Spain), Bram Bekaert (Belgium), Ewelina Pospiech (Poland), Arwin Ralf (the Netherlands), Ulises Toscanini (Argentina), Martin Bodner (Austria), Adrian Linacre (Australia), Weibo Liang (China), Kristiaan van der Gaag (Netherlands), Eduardo Avila (Brazil), Daniel Kling (Norway), Tom Callaghan (USA), Roland van Oorschot (Australia), Arnoud Kal (Netherlands), Thore Egeland (Norway), Nadia Pinto (Portugal), Federica Alessandrini (Italy), Ingo Bastisch (Germany), Katja Anslinger (Germany), Felix Bittner (the Netherlands), Nathan Scudder (Australia) and Tacha Hicks (Switzerland). The Scientific Program Committee expresses appreciation to the members of these selection committees and congratulations to the award winners.

At the ISFG General Assembly, Roland van Oorschot, from the Victoria Police Forensic Services Department (Australia), received the 2024 Scientific Prize for his global leadership in our understanding of DNA transfer, persistence, prevalence and recovery, beginning with his 1997 milestone Nature article "DNA fingerprints from fingerprints". Ángel Carracedo, from the University of Santiago de Compostela (Spain), received the ISFG Lifetime Achievement Award for being a leading contributor to our field for the past 40 years and pioneering many aspects of DNA analysis, sharing his knowledge and enthusiasm with the forensic genetics' community, from fellow experts to those just starting their careers. These prize winners were decided by the ISFG Board following candidate proposals submitted

by ISFG members and input from a panel of former prize winners (not including John Butler and Walther Parson who are on the ISFG Board). Roland has been invited to give the opening scientific prize plenary lecture at the next ISFG Congress in Montreal, Canada, to be held August 17-21, 2026.

As part of the ISFG General Assembly on September 11, 2024, Marielle Vennemann was elected as President, Walther Parson as Secretary, Lourdes Prieto as Treasurer, Corina Benschop was reelected Representative for Training and Education and Hwan Young Lee, from the Department of Forensic Medicine of Seoul National University, South Korea, was elected Representative of the Working Parties. Walther Parson was also elected Operational Manager for ISFG Interests. We are grateful for Sascha Willuweit who acted in this role following the passing of Peter Schneider in September 2022. In addition, during the General Assembly, honorary ISFG memberships were awarded to Antonio Alonso, Peter Gill, Lutz Roewer and Peter Schneider. We deeply appreciate Leonor Gusmão who diligently served on the ISFG Board from 2007 through 2024 as Treasurer, Representative of the Working Parties and Acting Secretary.

These published proceedings, with 165 articles, represents only a fraction of the knowledge shared during the week we spent together in Santiago de Compostela. We express our gratitude for the planners and participants who made ISFG 2024 a wonderful success!

John M. Butler*
on behalf of the ISFG Board
and Scientific Program Committee

* ISFG President; National Institute of Standards and Technology, Special Programs Office, 100 Bureau Drive, Mail Stop 6021, Gaithersburg, MD 20899-6021, USA; john.butler@nist.gov

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 27-31
DOI: https://dx.doi.org/10.15304/cc.2025.1869

NIST Research Grade Test Materials: A New Collaborative Effort to Address Measurement Challenges

Peter M. Vallone[1*], Lisa A. Borsuk[1], Carolyn R. Steffen[1], Sarah Riman[1], Kevin M. Kiesler[1], Erica L. Romsos[1]

[1]*U.S. National Institute of Standards and Technology, NIST 100 Bureau Drive, Gaithersburg, MD 20899-8314, USA*

[*]*Corresponding author at: National Institute of Standards and Technology, NIST 100 Bureau Drive, Gaithersburg, MD 20899-8314, USA, peter.vallone@nist.gov*

Abstract

As DNA typing technologies have advanced and sensitivity improved, measurement and interpretation issues related to low template, degraded, and complex DNA mixtures encountered in forensic laboratories have evolved. In turn, the forensic DNA typing community has requested more relevant reference materials to support the validation of their methods. The stochastic nature of these measurements, in tandem with the goal of reproducibly providing "casework-like" sample types, poses a challenge in meeting the robust characterization and the long-term stability required for a Standard Reference Material classification. NIST has introduced a new category of exploratory material called a Research Grade Test Material (RGTM) to collaboratively evaluate fit-for-purpose needs within a community. Laboratories can request the material with an agreement that data collected from the samples will be shared with NIST. A data portal hosted through STRBase (https://strbase.nist.gov/Information/RGTM_10235) was created to receive and publicly share results. The goal is that these samples and companion data portal resources will support validation efforts, facilitate data sharing, and guide further development of forensically relevant reference materials.

Keywords

Reference material, quantification, mixture, short tandem repeat, degraded DNA

1. Introduction

To address sample preparation and characterization challenges for reference materials, the NIST Office of Reference Materials has introduced a new classification of exploratory material called a Research Grade Test Material (RGTM). RGTMs aim to assess fit-for-purpose requirements collaboratively, enabling a more effective approach to complex measurement challenges [1]. Data from RGTMs may be generated collaboratively with customers to determine relevant values for what is being measured and the long-term stability of the material. RGTM 10235: Forensic DNA Typing Resource Samples was released to the forensic DNA community in August 2023. RGTM 10235 is composed of eight well-quantified DNA extracts. Samples include three single-source samples, two degraded samples, and three mixture samples.

2. Materials and methods

Samples 1, 3, and 5 (single source) were prepared at approximately 5 ng/µL in TE-4 buffer. Samples 2 and 4 (single source-degraded) were treated with UV light to induce DNA degradation. All samples except 3 and 4 were diluted in a TE-4 buffer containing 50 ng/µL of yeast tRNA (Thermo Fisher Scientific) to improve nucleic acid stability. Samples 6, 7, and 8 of the RGTM 10235 are DNA mixtures. Sample 6 is a two-person female-male mixture (90:10), Sample 7 is a three-person female-male-male mixture (20:20:60), and Sample 8 is a three-person female-male-male mixture (10:30:60). The components of the mixtures were quantified using digital PCR [2] and mixed in the appropriate volumes to create the intended mixture ratios. The primary landing page for RGTM 10235 is https://strbase.nist.gov/Information/RGTM_10235. Here, further information about the samples is hosted, as well as links to data submission instructions and data collected to date.

3. Results and Discussion

Figure 1 below is a composite graphic from the RGTM 10235 web pages. Panel A provides details on the materials and how to place an order. Additional information on the informed consent for the materials and specific NIST Terms and Conditions can be downloaded for review. Panel B lists the drop-down menu choices under

"NIST Resources" and "RGTM 10235". From these pages, the user can submit and view previously submitted data related to quantification measurements, STR typing allele calls, mixture interpretation, and mitochondrial sequencing. Panel C is an example plot of user-submitted DNA quantification measurements. Each concentration data point corresponds to the anonymous user number and the qPCR kit used to quantify the sample. An Excel file containing the full qPCR analysis conditions (e.g., instrument, software, calibration curve sample) can be downloaded from this page. Panel D is an example plot of the log LR values for Sample 6 (2-person mixture). The box plot represents the results of replicate PCR amplifications (n =3), followed by interpretation using probabilistic genotyping and subsequent assignment of likelihood ratio values. These experiments and interpretations were carried out over three separate time points (August 2023, December 2023, and May 2024). The STR typing conditions and mixture interpretation parameters used when assigning the likelihood ratios are provided on the web page. An external laboratory (Bode Technology Group, Lorton, VA) conducted the mixture analyses.

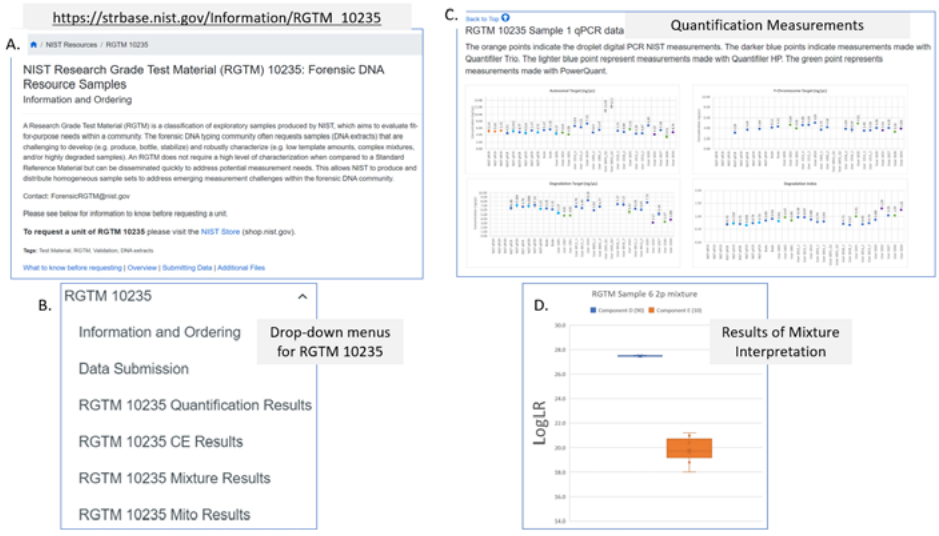


Figure 1. Example information and data available on the RGTM 10235 web pages (https://strbase.nist.gov/).

4. Conclusions

RGTM 10235 aims to offer the community a set of valuable samples that can be used in forensic DNA laboratories and compared anonymously to other practitioner datasets. The STRBase data portal enables the collection of externally performed measurements and subsequent posting of the results for various properties and characterization methods. The data from RGTM 10235 indicates that the material is stable and can help support forensic validation studies and training exercises. Additionally, RGTM 10235 provides longitudinal data for NIST to develop future reference material production better to keep current with the changing needs of the forensic DNA community.

5. Role of funding

This project was supported by the NIST Special Programs Office: *Forensic Genetics*.

6. Acknowledgements

Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. Certain commercial software, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

All work performed at NIST has been reviewed and approved by the National Institute of Standards and Technology Research Protections Office. This study was determined to be "not human subjects research" (often referred to as research not involving human subjects) as defined in U. S. Department of Commerce Regulations, 15 CFR 27, also known as the Common Rule (45 CFR 46, Subpart A), for the Protection of Human Subjects by the NIST Human Research Protections Office and therefore not subject to oversight by the NIST Institutional Review Board.

7. Conflict of interest

None.

8. References

[1] Beauchamp, C.R. et al., (2021) NIST Special Publication 260-136 2021 Edition Metrological Tools for the Reference Materials and Reference Instruments of the NIST Material Measurement Laboratory https://doi.org/10.6028/NIST.SP.260-136-2021 (Last accessed 8/28/2024)

[1] Romsos, E.L. et al., (2018) NIST Special Publication 260-189 Certification of Standard Reference Material® 2372a Human DNA Quantitation Standard. https://doi.org/10.6028/NIST.SP.260-189 (Last accessed 8/28/2024)

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 33-39
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Micromanipulated single cell subsampling, genetic analysis, and probabilistic genotyping for complex mixture deconvolution

Kaitlin Huffman[1], Erin Hanson[2,3], Jack Ballantyne[2,,3*]

[1]*Federal Bureau of Investigation Laboratory, DNA Support Unit, 2501 Investigation Parkway, Quantico, VA 22135, USA*

[2]*Department of Chemistry, University of Central Florida, Orlando, FL 32816-2366, USA*

[3]*National Center for Forensic Science, Orlando, FL 32816-2367, USA*

** Corresponding author at: jack.ballantyne@ucf.edu (407)-823-4440*

Abstract

Crime scene samples are often comprised of DNA mixtures from both victim(s) and perpetrator(s) to a crime. These mixtures can be particularly difficult to interpret due to overlapping alleles from multiple contributors or related individuals in addition to low templates of either the entire mixed sample or specific donors to the mixture. Due to the difficulties associated with complex mixture deconvolution, the ability to deconvolute DNA mixtures into their individual donor genotypes would significantly increase the number of cases in which probative results are obtained. In the present work, a single cell method referred to as direct single cell subsampling (DSCS) was used to collect single cells (or small cell subsets) from some of the most complex mixture scenarios (e.g., equimolar 3+ person mixtures, 4-person 1st degree relative mixtures, and mixtures with extreme minor donors). This resulted in a probative increase of information compared to standard methods.

Keywords

Single cell analysis; mixture deconvolution; probabilistic genotyping.

Introduction

Standard DNA mixture analysis focuses on mixture deconvolution at the end of the DNA workflow (i.e., extraction→ quantification→ amplification→ capillary electrophoresis→ mixture deconvolution). This can become particularly difficult

as the number of contributors (NOC) or the relatedness of contributors increases due to allele sharing often resulting in limited information in the most complex scenarios. In contrast, single cell analysis aims to decrease analysis complexity by deconvoluting mixtures up front prior to sample extraction [1–7]. Although low template in nature (i.e., ~6.6 pg for diploid cells and ~3.3 pg for haploid cells), single cell analysis can result in highly probative single source DNA genotypes. Genotypes attributed to the same donor can be utilized to either determine a database uploadable DNA profile via DBLR [6] or used with a PG systems' replicate analysis function [2–4,7].

Materials and methods

a. Mixture creation

Buccal swabs were agitated in 300 µL *of DNA suspension buffer. Samples were centrifuged at 300 RCF to create an epithelial cell pellet. The resulting supernatant was discarded. A cell suspension was created by adding 300* µL *of fresh DNA suspension buffer to each cell pellet. Cell concentrations were determined using the* Countess™ II FL (ThermoFisher Scientific, Carlsbad, CA, USA) automated cell counter. Various cell mixtures were created by adding the appropriate donor cell volumes to create the desired mixture ratio (e.g., equimolar 2-6 person mixtures). Cell suspensions/mixtures were stored at 4 °C.

b.Single cell analysis (DSCS)

A portion of each cell mixture (60 µL) was pipetted onto a Gel-Film® microscope slide and spread with a sterile cotton swab. Trypan blue solution was used to stain each mixture slide (1-2 mins). Slides were air-dried overnight. The Leica M205C stereomicroscope (190-240X magnification) was used for cell visualization and collection. A tungsten needle and 3MTM water-soluble adhesive was used to transfer individual or small cell subsamples (e.g., 1-5 cells) to sterile 0.2 mL PCR flat-cap tubes containing lysis solution. Detailed directions on slide creation and the cell collection method have previously been reported in an open access video demonstration [8]. Various lysis solutions were utilized (e.g., 5 µL PunchSolutionTM (Promega, Madison, WI, USA), 1 µL Prep-n-GoTM lysis buffer, or 1 µL Casework Direct Solution with 0.025 µL of a 50X

dilution of 1- thioglycerol). PunchSolutionTM samples were incubated with open tubes for lysis evaporation using a protocol of 70 °C→ 30 min. Prep-n-GoTM lysis samples were incubated at 90oC→ 20 mins; 25oC→ 15 mins. Casework Direct lysed samples were incubated at 70 °C→ 10 mins. After lysis, samples were amplified using GlobalFilerTM Express with a decreased reaction volume (i.e., 2 µL PCR mix, 2 µL primer mix, and 1 µL 5X AmpSolutionTM *(Promega, Madison, WI, USA)). Samples were amplified at 95oC→ 1 min; 32 cycles: 94oC→ 3 sec, 60oC→ 30 sec; 60oC→ 8 mins; 4oC→ hold. *Note: 5X AmpSolutionTM was not added to samples lysed with Prep-n-GoTM buffer. PG Software STRmix™ v2.8 was used to analyze each resulting subsample as either a single source or two-person mixture when appropriate. The FBI Caucasian allele frequency database was used, and the sub-source log (LR) reported. Replicate analysis was conducted on subsamples that returned inclusionary LRs (i.e., log(LR) > 1) for the same donor.

c. 'Bulk' analysis

A portion of each cell mixture (60 µL) was extracted using the AutoMate Express™ Forensic DNA Extraction System and PrepFiler Express™ Forensic DNA Extraction Kit (ThermoFisher Scientific, Carlsbad, CA, USA). Quantification was conducted using QuantifilerTM Trio (ThermoFisher Scientific, Carlsbad, CA, USA). The GlobalFilerTM amplification kit was used to amplify 1 ng of sample DNA at 29 cycles according to the manufacturers recommended protocol. One microliter of amplified product was added to 0.5 µL GeneScan™ 600 LIZ® size standard (ThermoFisher Scientific, Carlsbad, CA, USA) and 9.5 µL Hi-Di™ formamide (ThermoFisher Scientific, Carlsbad, CA, USA) and injected on an Applied Biosystems' 3500 Genetic Analyzer using Module J6 (15 s injection, 1.2 kV, 60 °C). Analysis was conducted with GeneMapper™ ID-X v1.6 (ThermoFisher Scientific, Carlsbad, CA, USA). PG Software STRmix™ v2.8 was used to analyze each resulting mixture according to the known NOC. The FBI Caucasian allele frequency database was used and the sub-source log (LR) reported.

d. DSCS workflow description

The DSCS workflow allows for subsampling from an original bulk mixture. Figure 1 provides an infographic of the DSCS method in which mini-mixtures are collected to improve the probative value obtainable from a

binary 1:50 mixture. Standard "bulk" sampling results in a single source profile of donor A while 5-cell subsamples can result in either single source profiles for donor A or 1:4 mini-mixed subsamples containing both individuals A and B which can subsequently be analyzed by PG. Most often, the DSCS method is utilized for collection of single cells resulting in single source profiles of all donors to a mixture.

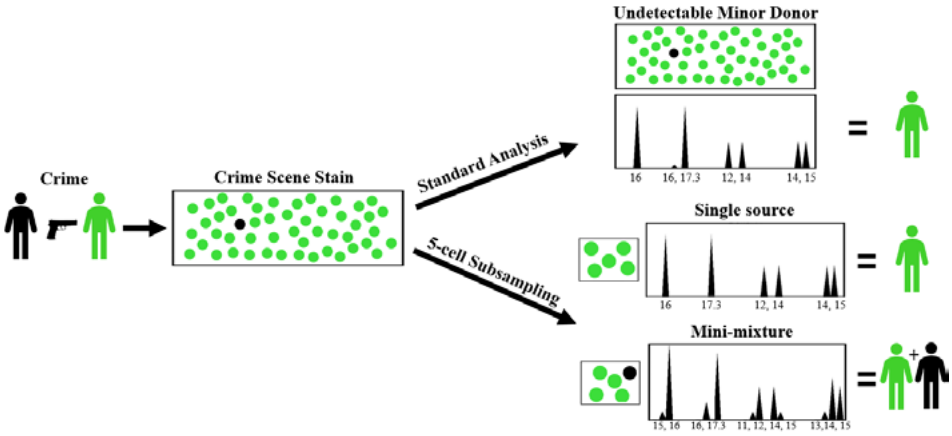


Figure 1. DSCS infographic showing the utility of mini-mixtures in analyzing a 1:50 bulk mixture.

Results

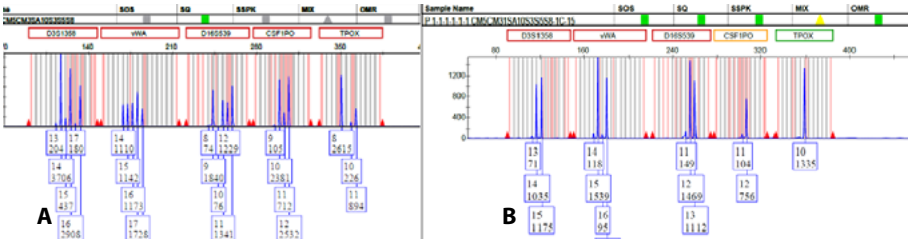


Figure 2. Single dye channel EPGs. A) 6-person equimolar mixture analyzed using the standard DNA workflow. B) Single cell analyzed using the DSCS workflow.

Table 1. Comparison of standard and DSCS STRmix™ results for various mixtures (M= mother; F= father; S= sibling; U= unrelated donor).

Mixture	Method	# of cells	POI 1 log(LR)	POI 2 log(LR)	POI 3 log(LR)	POI 4 log(LR)	POI 5 log(LR)	POI 6 log(LR)
Equimolar 2p [2]	Standard		17	16				
	DSCS[a]	20x 1C & 2C	27	27				
Equimolar 3p [2]	Standard		25	13	9			
	DSCS[a]	20x 1C & 2C	28	27	22			
Equimolar 4p [2]	Standard		7	15	8	4		
	DSCS[a]	30x 1C & 2C	22	25	20	17		
Equimolar 5p [2]	Standard		3	4	9	8	3	
	DSCS[a]	40x 1C & 2C	24	27	25	13	20	
Equimolar 6p [2,7]	Standard		8	8	1	8	4	5
	DSCS[a]	30x 1C & 2C	25	26	20	28	17	8
	DSCS	30x 1C	27	27	22	14	25	25
MFUU [3]	Standard		23	11	5	9	6b	2[b]
	DSCS	40x 1C & 2C	25	25	17	27	0[b]	0[b]
MFCU [3]	Standard		10	6	11	8	6[b]	6[b]
	DSCS	40x 1C & 2C	26	25	16	26	0[b]	0[b]
SSSS [3]	Standard		10	13	12	12	8[b]	
	DSCS	40x 1C	22	24	23	24	0[b]	
1:50 [6,9]	Standard		1	27				
	DSCS	75x 5C	11	28				
	DSCS[c]		23	NA				

[a] Prior to lysis improvement

[b] Non-donor relative

[c] DBLRTM common donor analysis

Discussion

The presented studies establish single cell methods to overcome some of the current limitations associated with DNA mixture analysis. By sampling individual cells, highly probative single source DNA genotypes are obtainable for donors to complex mixtures (Figure 2) which may otherwise be masked due to allele sharing (e.g., high NOCs or related donors) or undetectable due to low templates (e.g., extreme minor donors). Significant gain of information due to DSCS analysis is exemplified in Table 1.

Additionally, microscopic verification of certain cell types (e.g., sperm) can definitively link the resulting genotype to a specific body fluid. Although not intended to supplant current bulk mixture methods, the addition of single cell analysis could further inform standard analysis (e.g., providing a single source DNA profile to condition the bulk mixture upon or further informing NOC estimations).

Conclusion

The addition of a single cell analysis workflow could increase the number of complex cases in which a POI is definitively implicated or excluded as a true donor.

Acknowledgments

The authors would like to thank the state of Florida for the initial seed funding for this Project and the anonymous donors who provided samples.

Conflict of interest statement

There were no competing interests to disclose. Names of commercial manufacturers are provided for identification purposes only, and inclusion does not imply endorsement of the manufacturer, or its products or services by the FBI. The views expressed are those of the authors and do not necessarily reflect the official policy or position of the FBI or the U.S. Government.

References

[1] K. Huffman, E. Hanson, J. Ballantyne, Recovery of single source DNA profiles from mixtures by direct single cell subsampling and simplified micromanipulation, Science & Justice 61 (2021) 13–25. https://doi.org/10.1016/j.scijus.2020.10.005.

[2] K. Huffman, E. Hanson, J. Ballantyne, Probabilistic genotyping of single cell replicates from complex DNA mixtures recovers higher contributor LRs than standard analysis, Science & Justice 62 (2022) 156–163. https://doi.org/10.1016/j.scijus.2022.01.003.

[3] K. Huffman, J. Ballantyne, Probabilistic Genotyping of Single Cell Replicates from Mixtures Involving First-Degree Relatives Prevents the False Inclusions of Non-Donor Relatives, Genes (Basel) 13 (2022) 1658. https://doi.org/10.3390/genes13091658.

[4] K. Huffman, J. Ballantyne, Validation of Probabilistic Genotyping Software for Single Cell STR Analysis, Genes (Basel) 14 (2023) 674. https://doi.org/10.3390/genes14030674.

[5] K. Huffman, E. Hanson, J. Ballantyne, Y-STR mixture deconvolution by single-cell analysis, J Forensic Sci 68 (2023) 275–288. https://doi.org/10.1111/1556-4029.15150.

[6] K. Huffman, M. Kruijver, J. Ballantyne, D. Taylor, Carrying out common DNA donor analysis using DBLRTM on two or five-cell mini-mixture subsamples for improved discrimination power in complex DNA mixtures, Forensic Sci Int Genet 66 (2023) 102908. https://doi.org/10.1016/j.fsigen.2023.102908.

[7] K. Huffman, J. Ballantyne, Single cell genomics applications in forensic science: Current state and future directions, IScience 26 (2023) 107961. https://doi.org/10.1016/j.isci.2023.107961.

[8] K. Farash, E.K. Hanson, J. Ballantyne, Enhanced Genetic Analysis of Single Human Bioparticles Recovered by Simplified Micromanipulation from Forensic "Touch DNA" Evidence, Journal of Visualized Experiments (2015). https://doi.org/10.3791/52612.

[9] K. Huffman, E. Hanson, J. Ballantyne, Cell Subsampling Recovers Probative DNA Profile Information from Unresolvable/Undetectable Minor Donors in Mixtures, Genes (Basel) 13 (2022) 1117. https://doi.org/10.3390/genes13071117.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 41-48
DOI: https://dx.doi.org/10.15304/cc.2025.1869

More Markers, Better Results? Predicting Evolutionary Y-Haplogroups R1b and I1 with NevGen using the Yfiler® Plus Kit versus YForGen Kit

Heleen Coreelman[1,2,3,*], Cleo Coeman[1,2], Ronny Decorte[2,4], Ellen Decaestecker[3], Sofie Claerhout[1,2]

[1] *Crime Scene Y-DNA investigation, Interdisciplinary Research Facility Life Sciences, Department of Biology, KU Leuven Kulak, Kortrijk, Belgium*

[2] *Forensic Biomedical Sciences, Department of Imaging and Pathology, KU Leuven, Leuven, Belgium*

[3] *MicrobiomeEcoEvo group, Interdisciplinary Research Facility Life Sciences, Department of Biology, KU Leuven Kulak, Kortrijk, Belgium*

[4] *Laboratory of Forensic Genetics, Department of Forensic Medicine, UZ Leuven, Leuven, Belgium*

** Corresponding author at: Crime Scene Y-DNA investigation, Interdisciplinary Research Facility Life Sciences, Department of Biology, KU Leuven Kulak, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium. E-mail address: heleen.coreelman@kuleuven.be (H. Coreelman).*

Abstract

The male Y-chromosome plays a pivotal role in population and forensic genetics. Its rather conserved paternal inheritance provides valuable insights into human ancestry, biogeographical origins and DNA kinship investigations. While slowly mutating Y-chromosomal single nucleotide polymorphisms (Y-SNPs) are traditionally used to assign evolutionary Y-haplogroups, the process is labour-intensive, time-consuming and costly. Therefore, there has been a growing interest in seeking other methods for Y-haplogroup assignment. An efficient alternative approach is related to the more rapid mutating familial Y-chromosomal short tandem repeats (Y-STRs). To this purpose, several online Y-STR haplogroup predictors have been developed. Nevertheless, there has been some doubt about their effectiveness for forensic applications. Here, the performance of the NevGen Haplogroup Predictor was explored for Y-subhaplogroups R1b and I1 using a different number of input Y-STR loci, namely 23 Y-STRs from the *Yfiler® Plus Kit (Thermo Fisher Scientific)* and 42 Y-STRs from our in-house developed YForGen Kit (UZ/KU Leuven). To illustrate the impact of input Y-STR variability on prediction accuracy in practice, two case studies are presented. Comparing the haplogroup predictions to the Y-SNP-derived (sub)haplogroup demonstrated that increasing the number of input Y-STRs enhances the prediction accu-

racy. These results underline NevGen's capability to produce higher-confidence predictions when more Y-STR markers are included, supporting its reliability to estimate a man's biogeographical origin on a broader phylogenetic level.

Keywords

Y-chromosome, Biogeographical origin estimation, Y-SNP haplogroups, NevGen Y-STR Haplogroup Predictor.

Introduction

With its strict paternal inheritance, the male-specific Y-chromosome plays a significant role in population and forensic genetics. Unlike autosomes, the Y-chromosome remains largely unchanged across generations, making it a powerful marker for tracing paternal lineages [1]. Single nucleotide polymorphisms on the Y-chromosome (Y-SNPs) evolve at a relatively slow rate (10-9-10-8 mutations per generation) [2], enabling the assignment of men to one of the twenty major evolutionary Y-haplogroups (A to T) and their more detailed Y-subhaplogroups [3]. These haplogroups are often used to infer a man's biogeographical origin [1]. Unfortunately, Y-SNP typing is both time-consuming and costly, presenting challenges in forensic contexts where rapid and cost-efficient methods are paramount. To address these limitations, researchers have explored the potential of using Y-chromosomal short tandem repeats (Y-STRs), which mutate at a faster rate (10-4-10-2 mutations per generation) than Y-SNPs [2], to predict Y-haplogroups. Y-STRs are widely used in forensic casework to establish familial relationships [1]. Due to the correlation between Y-haplotypes and specific Y-haplogroups, these faster-evolving markers could serve as a cost-effective alternative to Y-SNP typing for haplogroup assignment. Therefore, several computational tools, such as Whit Athey's [4], NevGen [5] and PredYMaLe [6] Haplogroup Predictors, have been developed for Y-haplogroup estimation based on Y-STR profiles.

While these Y-STR-based predictors offer potential advantages in terms of speed and efficiency, questions remain about their accuracy in forensic applications. Our study aims to evaluate the prediction accuracy of different Y-STR haplogroup predictors in our large CSY-dataset including men from Belgium and the Netherlands. Moreover, our study aims to evaluate the impact of input Y-STR variability on prediction accuracy using Y-STR markers included in the commercially available Yfiler® Plus Kit (Thermo Fisher Scientific; 27 Y-STRs) versus our in-house developed YForGen Kit (UZ/KU Leuven; 46 Y-STRs including all Y-STRs from the Yfiler® Plus Kit) [7]. The increase of Y-STR markers in the YForGen Kit

can potentially offer more accurate haplogroup predictions. In this study, the prediction accuracy of the NevGen Haplogroup Predictor was investigated for Y-subhaplogroups R1b and I1 using a different number of input Y-STR loci. NevGen was selected as Bayesian-allele-frequency approach for its capacity to predict a wider range of haplogroups and its ability to account for correlations between STR values compared to the Whit Athey Haplogroup Predictor [5]. Additionally, two case examples were selected to demonstrate the impact of including more Y-STR markers on NevGen's haplogroup prediction accuracy. With this, we aim to provide insights to improve the reliability of Y-haplogroup assignment, thereby enhancing the utility of Y-STR data in both population genetics and forensic investigations.

Materials and methods

Database

Our CSY-database includes 2,476 males from Belgium and the Netherlands [7]. Y-DNA analysis and publication was approved by the Ethics Committee Research of University Hospitals Leuven / KU Leuven (S55864, S59085) and written informed consent was obtained from all participants. DNA samples were collected, extracted and genotyped using capillary electrophoresis by 81 Y-SNPs for (sub)haplogroup assignment and 38 to 46 Y-STRs for haplotype assignment. Y-DNA data is available through Y-STR Haplotype Reference Database (YHRD, https://yhrd.org/) under accession numbers YA003651, YA003652, YA003653, YA003739, YA003740, YA003741, YA003742, YA004300 and YA004301.

Y-STR haplogroup predictions

The NevGen Haplogroup Predictor (https://www.nevgen.org/) was used for Y-(sub)haplogroup assignment. This online tool uses a Bayesian-allele-frequency approach in predicting the (sub)haplogroup to which a Y-STR haplotype belongs to. Furthermore, it uses correlations between values of Y-STR pairs during the calculation of probabilities. NevGen input requires a string of Y-STR values in FamilyTreeDNA order, separated by commas. The obtained output is the predicted Y-(sub)haplogroup(s) supported by the online software tool, their probability score(s) as well as fitness score(s). For each man, the Y-(sub)haplogroup predictions were conducted twice: once with 23 Y-STR loci included in the Yfiler® Plus Kit (Thermo Fisher Scientific; 4 Y-STR loci are not supported by NevGen) and once with 42 Y-STR loci included in the YForGen Kit (*UZ/KU Leuven;* 4 Y-STR loci are not supported by NevGen*)*. The Y-(sub)haplogroup prediction with the highest

probability and fitness score was used for further analyses. When the highest prediction probability was assigned to an unsupported subclade, the next Y-(sub)haplogroup was selected.

Prediction accuracy

The prediction accuracy was investigated in terms of Y-(sub)haplogroup and number of input Y-STR loci. For each man in our CSY-database belonging to Y-subhaplogroup R1b or I1, the prediction accuracy was assessed by comparing NevGen's Y-(sub)haplogroup prediction with the Y-SNP-derived (sub)haplogroup at various depth levels in the phylogenetic tree. Predictions that did not meet a specific depth level were excluded from the corresponding analysis, e.g. if the prediction identified Y-haplogroup T, it was only considered at the first depth level.

Results

Figure 1 shows the Y-haplogroup distribution of our CSY-database, consisting of haplogroups R (62%), I (22%), E (6%), J (5%), G (4%), and T, Q, L, H, N (1%). This distribution is representative of the Belgian and Dutch population [8].

Y-HG	Relative frequency (%)	
	CSY-database	BE and NL
R	61.75	62.67
I	21.65	25.79
E	5.65	3.29
J	5.53	3.85
G	4.12	2.94
T	0.53	0.63
Q	0.36	0.31
L	0.28	0.14
H	0.08	0.03
N	0.04	0.03

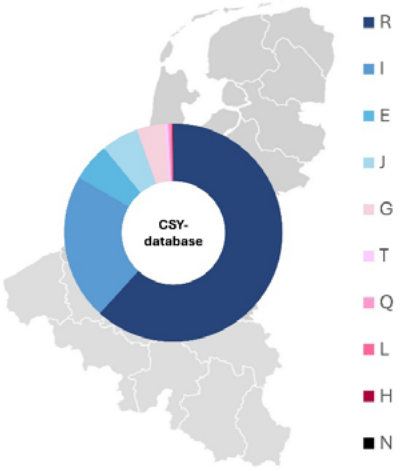
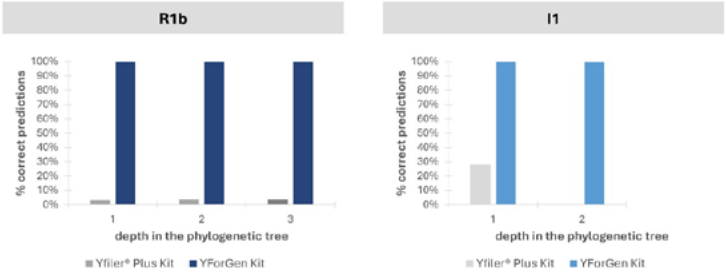


Figure 1. Y-haplogroup (Y-HG) distribution of our CSY-database including 2,476 men from Belgium and the Netherlands [7]. Their Y-haplogroup was defined via 81 Y-SNPs. The database is representative of the Belgian (BE) and Dutch (NL) population [8].

As the two most common Y-subhaplogroups in our dataset are R1b and I1, with respectively 1,428 and 319 men, the prediction accuracy within these groups was investigated.

Figure 2 demonstrates that NevGen's haplogroup predictions became significantly more accurate with increased input data ($p<0.05$). For Y-subhaplogroup R1b, the proportion of correct predictions increased from 3.29% (47/1428) to 100% (1428/1428) at the first depth level (R), from 3.74% (47/1258) to 100% (1428/1428) at the second level (R1), and from 3.74% (47/1258) to 99.93% (1427/1428) at the third level (R1b). Similarly, for Y-subhaplogroup I1, the proportion of correct predictions improved from 27.90% (89/319) to 100% (319/319) at the first depth level (I) and from 0.32% (1/315) to 99.69% (318/319) at the second level (I1). This improvement in accuracy is further illustrated by two case examples. In the case of Man A, the Y-subhaplogroup is R1b1a1b1a1a2a1b1a1. When using the input data of the Yfiler® Plus Kit, NevGen incorrectly predicted the subhaplogroup as I2a1. However, with the input data of the YForGen Kit, the prediction shifted to R1b, which correctly identifies the broader subhaplogroup. In the case of Man B, the Y-subhaplogroup is I1a1b1a1. By using the input data of the Yfiler® Plus Kit, NevGen incorrectly predicted J2a1 as (sub)haplogroup. When the analysis was repeated with input data of the YForGen Kit, NevGen correctly predicted the subhaplogroup as I1, which corresponds to the broader subhaplogroup.



Man A		Y-(sub)HG
Y-SNP analysis		R1b1a1b1a1a2a1b1a1
Prediction	Yfiler® Plus Kit	I2a1
	YForGen Kit	R1b

Man B		Y-(sub)HG
Y-SNP analysis		I1a1b1a1
Prediction	Yfiler® Plus Kit	J2a1
	YForGen Kit	I1

Figure 2. *Prediction accuracy of the NevGen Haplogroup Predictor for Y-subhaplogroups (Y-subHG) R1b (***left***) and I1 (***right***), based on Y-STR data from our CSY-database. The men's Y-(sub)HG and Y-haplotype were characterised via Y-SNPs and Y-STRs, respectively. Input data for the NevGen Haplogroup Predictor were either 23 Y-STR loci included in the Yfiler® Plus Kit or 42 Y-STR loci included in the YForGen Kit. The proportion of correct predictions per depth in the phylogenetic tree is indicated in the upper graph. Two case examples are elaborated in the lower table.*

Discussion

The significant improvements in prediction accuracy for Y-subhaplogroups R1b and I1 when moving from 23 to 42 Y-STR markers shows that additional markers provide crucial data that improve Y-(sub)haplogroup prediction accuracy. Using the Y-STR loci from the Yfiler® Plus Kit, predictions were mostly incorrect and misaligned with the Y-SNP-derived (sub)haplogroups. These incorrect predictions reflect the limitations of using smaller marker sets, which could be attributed to the fact that fewer markers may not capture the full genetic diversity necessary for accurate Y-(sub)haplogroup estimation. It has been demonstrated that the similarity of Y-STR haplotypes across different Y-(sub)haplogroups might reduce the reliability of Y-(sub)haplogroup predictions [9]. The shift to Y-STR loci from the YForGen Kit resulted in more accurate predictions of broader Y-(sub)haplogroups, demonstrating that increasing the number of input Y-STRs, including rapidly mutating ones, improves prediction accuracy. As the number of markers increases, the Bayesian-allele-frequency approach has more data to work with, provided that those Y-STRs are available for input, leading to more refined and accurate predictions. Importantly, a deeper subclade identification has not yet been conducted, despite the predicted final Y-SNP being available in NevGen's output. For example, according to ISOGG Y-DNA Haplogroup Tree 2019-2020, Y-SNP M269 corresponds to Y-subhaplogroup R1b1a1b. Including this information could be explored in future research, as it would provide insights not only into a man's overarching Y-(sub) haplogroup but also into his deeper phylogenetic subhaplogroup, determining the extent to which NevGen can resolve deeper subclades based solely on Y-STR data.

The findings suggest that for high accuracy with the NevGen Haplogroup Predictor, more comprehensive Y-STR marker sets such as the YForGen Kit are essential. Especially in forensic applications, where accurate haplogroup assignment can offer deeper insights into biogeographical origins. The results support NevGen's utility as a predictive tool in forensics, provided that it is used with an extensive marker set. Nevertheless, confirmation and further refinement based on Y-SNP sequencing might still be desirable to reach the most accurate Y-haplogroup information on a deeper subclade level.

Conclusion

The evidence from our study highlights the value of using more Y-STR markers for R1b and I1 haplogroup predictions with the NevGen Haplogroup Predictor. Moving

forward, investigators should give priority to the inclusion of larger and more comprehensive Y-STR marker sets than the Yfiler® Plus Kit to exclude false conclusions and ensure the highest possible prediction accuracy in forensic applications.

Acknowledgments

We extend our gratitude to all DNA donors for their valuable participation in our previous genetic genealogy projects. This research was funded by FWO (application numbers: H. Coreelman 1154123N and S. Claerhout 1265622N).

Conflict of interest statement

The authors declare that there is no conflict of interest.

References

[1] M. Kayser, Forensic use of Y-chromosome DNA: a general overview, Hum Genet 136 (2017) 621–635.

[2] O. Balanovsky, Toward a consensus on SNP and STR mutation rates on the human Y-chromosome, Hum Genet 136 (2017) 575–590.

[3] M. Van Oven, A. Van Geystelen, M. Kayser, et al., Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome, Hum Mutat 35 (2014) 187–191.

[4] T.W. Athey, Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach, Journal of Genetic Genealogy 2 (2006) 34–39.

[5] About NevGen haplogroup predictor, (n.d.). https://www.nevgen.org/AboutNevGen.html (accessed September 16, 2024).

[6] C. Bouakaze, F. Delehelle, N. Saenz-Oyhéréguy, et al., Predicting haplogroups using a versatile machine learning program (PredYMaLe) on a new mutationally balanced 32 Y-STR multiplex (CombYplex): Unlocking the full potential of the human STR mutation rate spectrum to estimate forensic parameters, Forensic Sci Int Genet 48 (2020) 102342.

[7] S. Claerhout, M. Vandenbosch, K. Nivelle, et al., Determining Y-STR mutation rates in deep-routing genealogies: Identification of haplogroup differences, Forensic Sci Int Genet 34 (2018) 1–10.

[8] E. Altena, R. Smeding, K.J. van der Gaag, et al., The Dutch Y-chromosomal landscape, European Journal of Human Genetics 28 (2020) 287.

[9] C.C. Wang, L.X. Wang, R. Shrestha, et al., Convergence of Y Chromosome STR Haplotypes from Different SNP Haplogroups Compromises Accuracy of Haplogroup Prediction, Journal of Genetics and Genomics 42 (2015) 403–407.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 49-56
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Predicting the LR value of the mixed DNA profile using random forest regression for database searching

Yuhan Hu[1,†], Xiaoyi Xu[2,†], Ji Zhang[1,*]

[1] *West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu, China*

[2] *School of Chemical Engineering, Sichuan University, Chengdu, China*

† *Yuhan Hu and Xiaoyi Xu contribute equally to the article.*

** Corresponding author at: West China School of Basic Medical Sciences & Forensic Medicine, zhangj@scu.edu.cn,*

Abstract

Comparison of multiple reference individuals of a DNA database with a mixed DNA profile from an evidentiary trace is an important approach for forensic investigation. Several search methods exist for accurately, but time-consumingly, identifying potential suspects through qualitative or quantitative LR calculations. In this study, the Random Forest Regression algorithm was used to analyze the mixed DNA profiles for rapid LR prediction.

The mixed DNA profiles of the PROVEDIt dataset, generated by five kits, with two to three contributors, varying injection times, and different assay platforms, were used to calculate the log10(LR) for each contributor separately by Euroformix. Kit-specific and injection time-specific AT values, as well as drop-in parameters, were determined using single-source profiles. Degradation and backward stutters were considered during computation. A total of 6,798 pieces of data were input into the algorithm. After feature selection, algorithm training, and testing, the model's performance was evaluated.

The optimal model utilized seven features, achieving a cross-validation score of 0.8611. The learning curves demonstrated strong model fit, with R^2 values of 0.9881 for the training set and 0.9017 for the test set. Furthermore, the differences between predicted and original Log10(LR) values followed a normal distribution, Fifty percent of the difference was concentrated in the range of [-1.189, 1.455], while ninety percent of the data fell within the interval of [-4.338, 4.108]. 99.6% of predicted values remaining positive when the original Log10(LR) was above zero.

The random forest regression model offers a rapid and efficient approach for predicting LRs in mixed DNA profiles, and demonstrates strong potential for database searching.

Keywords
Database searching, Machine learning, Likelihood ratio, DNA mixture

Introduction

Comparing multiple reference individuals from a DNA database against a mixed DNA profile obtained from evidentiary traces is a crucial method in forensic investigations. Utilizing the likelihood ratio (LR) as a match indicator is an effective strategy for ranking candidates above a specific threshold, thereby aiding in the identification of potential target individuals [1].

To enhance the efficiency of this process, various software tools have been developed to exclude irrelevant individuals and narrow the pool of potential contributors [2-5]. While these software programs have demonstrated significant advantages in improving detection efficiency and accuracy, they still face challenges related to computational speed and the risk of incorrectly including unrelated individuals, especially when working with large databases containing millions or even tens of millions of records.

In this study, we propose using the LR as a label and calculating various features derived from each candidate profile to establish a relationship between these features and the label. This approach leverages regression algorithms in machine learning, which can significantly expedite the searching process once an effective model is established.

Method

LR calculation

In this study, 2-3 person's mixture profile was analyzed in the PROVEDIt dataset [6], which generated using four different commercial kits under various laboratory conditions. The Filtered version of the dataset was selected, which excluded artifacts such as pull-up, minus-A, and -2bp stutter of SE33. Kit-specific and injection time-specific analytical thresholds (AT) was determined using single-source profiles, and drop-in parameter embedded in the gamma model was estimated based on the method described in [7].

The LR was calculated by using Euroformix software [8], with allele frequencies from the NIST1036 Caucasian population. The degradation and

backward stutter models were applied, while the forward stutter model was disabled. The Log10(LR) value served as the label for each data point. The mutual exclusion assumptions were set as follows:

H_p: *POI + (N-1) UN*

H_d: *(N) UN*

Where the person of Interest (POI) refers to one of the true contributors of the mixture profile, UN represents unrelated individuals, and N denotes the number of contributors.

Feature Selection

Several features derived from the DNA profile were calculated, including both qualitative and quantitative features. The qualitative features were based on the comparison between the number of alleles in the reference and mixture profiles, while the quantitative features were related to the peak heights in the profile. Further details of these features are provided in Table 1, where the expectation, coefficient of variation, and degradation coefficient of the peak heights were estimated using a gamma distribution model, based on the averaged peak heights at each locus.

Table 1. *Features used to construct the machine learning algorithm.*

	Feature	*Description* [a]
qualitative features	*AlleleSum*	*Total number of alleles in E*
	MAC	*Matching allele count [9]*
	NoDropout Ratio	*MAC / Total number of alleles in R*
quantitative features	*PHsum*	*Sum of peak heights in E*
	MAPmax	*Maximum peak height of alleles matching R*
	MAPmin	*Minimum peak height of alleles matching R*
	MAPratio	*Sum of peak heights of alleles matching R/ PHsum*
	PHmu	*Peak height expectation*
	PHomgea	*Peak height coefficient of variation*
	PHdeg	*Peak height degradation coefficient*

[a] *E refers to DNA mixture profile, R refers to POI reference profile.*

Random Forest Model Construction

Following the calculation of LR and feature extraction, a random forest regression model was constructed. During feature engineering, the features were

standardized using Z-Score normalization, and the dataset was split into 80% for training and 20% for testing. After conducting feature importance analysis and applying recursive feature elimination, the final set of features for the model was selected. In the model training phase, grid search with 5-fold cross-validation was employed to optimize hyperparameters. The model's performance was evaluated using the coefficient of determination (R^2), where a higher R^2 value indicates a better model fit.

Results

Feature Engineering

A total of 6,798 data points, each with features and corresponding labels, were available for constructing the regression model. The feature importance is illustrated in Fig. 1. Using Recursive Feature Elimination (RFE), features were systematically removed one at a time, starting with the least important. This process revealed that the optimal number of features for the model was 7, resulting in a best cross-validation score of 0.8611.

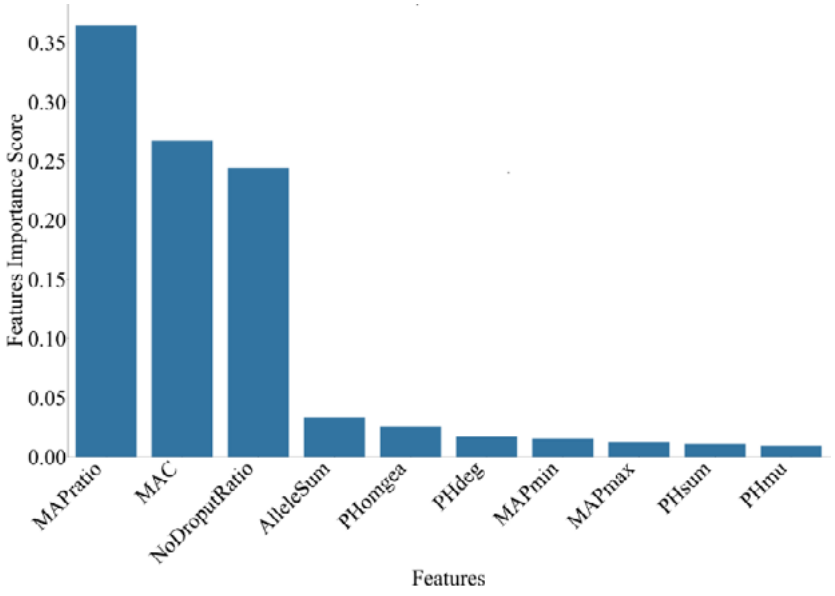


Fig. 1. *The feature importance scores for candidate features.*

Model Performance

In cross-validation, learning curves demonstrate the accuracy of both the training and validation sets as the amount of training data increases. As shown in Fig. 2, the accuracy of the training set remains consistently high, while the accuracy of the validation set gradually improves, with both curves converging over time. The R^2 for the training set was 0.9881, and for the test set, it was 0.9017, indicating a strong model fit.

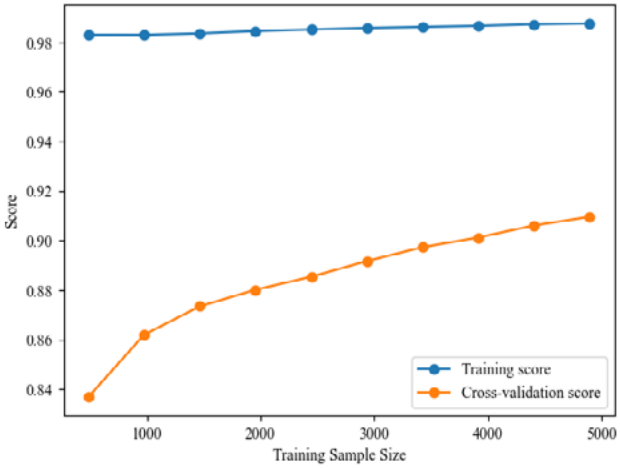


Fig. 2. *Learning curve assessing the random forest algorithm.*

Result of Test Set

The differences between the predicted and original Log10(LR) are shown in Fig. 3. The value in the test set follows a normal distribution, with a mean of 0.06. Fifty percent of these differences are concentrated within the range of [-1.189, 1.455], while ninety percent fall within the interval of [-4.338, 4.108]. Among the test samples where the original Log10(LR) was above zero, 99.6% of the predicted values were also greater than zero, supporting the prosecutor's hypothesis. Only five instances had predicted values below zero, specifically -1.061, -0.808, -0.770, -0.756, and -0.301, which corresponded to original values of 0.0422, 0.9560, 0.1701, 1.3440, and 0.5511, respectively.

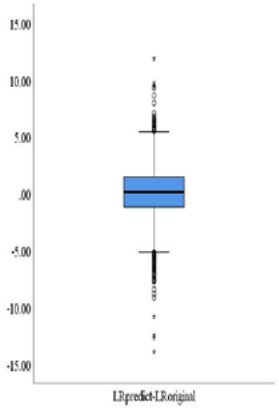


Fig. 3. The box plot of differences between predicted and original Log10(LR)

Discussion

In this study, we established a regression relationship between the LR and the features derived from the DNA profiles. During this process, we selected features relevant to LR calculation, with the most significant being the MAP ratio, MAC, and NoDropout Ratio. These features indicate the proportion of contributor in the mixture profile play a critical role in the model. They are most relevant factors influencing the LR value of true contributors.

Although the difference between the predicted and original Log10(LR) values can exhibit extreme variations—such as the largest absolute difference of 13.8806—the model demonstrates excellent categorization. Specifically, the majority of predicted values remain positive for reference individuals with positive original values. However, original values that are approximately close to 0 tend to shift the predicted categories into negative values. Further data are needed to investigate the impact of original values on predicted outcomes.

Furthermore, due to the significantly high number of non-contributors in the database, we need to include more non-contributors in the training and test sets in the future to evaluate the actual efficacy of random forest regression in database searching.

Conclusion

The Random Forest Regression offers a rapid approach for predicting the LRs of multiple suspects within mixed DNA profiles, aiding in the evaluation of a potential suspect's contribution to an evidence profile. Regression algorithms may provide an efficient and streamlined way to target suspects in large databases for future forensic investigations.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (82230064, 82072120).

Conflict of interest statement

The authors declare no conflict of interest.

References

[1] C.C.G. Benschop, J.d. Jong, L.v.d. Merwe, H. Haned, Adapting a likelihood ratio model to enable searching DNA databases with complex STR DNA profiles, 2016.

[2] Ø. Bleka, L. Prieto, P. Gill, CaseSolver: An investigative open source expert system based on EuroForMix, Forensic Science International: Genetics. 41 (2019) 83-92, https://doi.org/https://doi.org/10.1016/j.fsigen.2019.03.024

[3] Ø. Bleka, M. Bouzga, A. Fonneløp, P. Gill, dnamatch2: An open source software to carry out large scale database searches of mixtures using qualitative and quantitative models, Forensic Science International: Genetics Supplement Series. 6 (2017) e404-e406, https://doi.org/https://doi.org/10.1016/j.fsigss.2017.09.160

[4] C.C.G. Benschop, L. van de Merwe, J. de Jong, V. Vanvooren, M. Kempenaers, C.P. Kees van der Beek, et al., Validation of SmartRank: A likelihood ratio software for searching national DNA databases with complex DNA profiles, Forensic science international. Genetics. 29 (2017) 145-153, https://doi.org/10.1016/j.fsigen.2017.04.008

[5] J. Hoogenboom, T. Sijen, C. Benschop, ProbRank: An efficient DNA database search method for complex mixtures per a quantitative likelihood ratio model, Forensic Science International: Genetics. 65 (2023) 102884, https://doi.org/https://doi.org/10.1016/j.fsigen.2023.102884

[6] S. Riman, H. Iyer, P.M. Vallone, Examining performance and likelihood ratios for two likelihood ratio systems using the PROVEDIt dataset, PloS one. 16 (9) (2021) e0256714, https://doi.org/10.1371/journal.pone.0256714

[7] Ø. Bleka, L. Prieto, P. Gill, EFMrep: An extension of EuroForMix for improved combination of STR DNA mixture profiles, Forensic science international. Genetics. 61 (2022) 102771, https://doi.org/10.1016/j.fsigen.2022.102771

[8] Ø. Bleka, G. Storvik, P. Gill, EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts, Forensic Science International: Genetics. 21 (2016) 35-44, https://doi.org/10.1016/j.fsigen.2015.11.008

[9] Ø. Bleka, G. Dørum, H. Haned, P. Gill, Database extraction strategies for low-template evidence, Forensic science international. Genetics. 9 (2014) 134-141, https://doi.org/10.1016/j.fsigen.2013.11.006

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 57-59
DOI: https://dx.doi.org/10.15304/cc.2025.1869

DNA Mixture evidence, and Baskerville data – what isn't there can exonerate

Charles Brenner

PhD mathematician. UC Berkeley Human Rights

Charles@DNA-VIEW.com

Abstract

Probabilistic mixture analysis generally defines an "analytic threshold" under which epg data is ignored as unreliable – possibly garbage (bad evidence). Ignoring the sub-threshold data is considered "conservative", meaning that although some of the sub-threshold data is legitimate evidence against the suspect, we prefer to discard that good evidence rather than risk convicting a suspect on bad evidence. However, in this report I argue that the apparent conservativeness is seriously bad logic for two distinct reasons.

1. Withholding questionable data is generally a favor when the suspect is a DNA contributor – i.e. guilty. But since withholding information is in general withholding truth, hiding sub-threshold data is no favor when the suspect is innocent. On the contrary, hiding data in general means hiding the truth hence tends to "frame" – that is, to convict or inculpate – the innocent.
2. The assumption that sub-threshold data is unreliable, is also not correct. The idea makes sense for visible data, but *lack* of any signal at an allele position – a void – is also evidence, and is reliable.

Key words

Sub-threshold, void, confirmation bias, guilty, innocent

Introduction

The main section of this paper presents two important, simple, and universally overlooked errors in evaluating crime scene DNA evidence. Both errors pertain to the use of an analytic threshold.

A recent criminal case provides an example.

For simplicity and convenience, herein "suspect" means suspected of having contributed DNA to crime scene evidence, and the words "guilty" and "innocent" merely mean whether or not the suspect contributed to the DNA evidence.

The situation of most interest for this discussion is when some alleles consistent with the suspect are found not much above the analytical threshold, so that dropout (of sister alleles) is likely if the suspect is guilty.

Two universal misconceptions about the analytic threshold idea

B.1. The error of supposing that the analytical threshold favors the suspect

Given the epg from crime scene DNA of one or more contributors and reference genotype of a suspect, and supposing the analyst by eyeball sees obvious "exclusion" – my impression is that laboratory analysts typically do not bother with computer calculation, which can take minutes or more. Consequently, when the epg surmounts the eyeball exclusion hurdle, human nature – confirmation bias – might lean the analyst toward thinking of the suspect as guilty, might fall into supposing that it is their job to prove it. For such analysts, hiding sub-threshold information is to ensure that, at the cost of a smaller LR, the suspect can't complain about questionable evidence.

My hypothesis of confirmation bias isn't intended as an argument, but merely as partial explanation for how we have come to the wrong-headed idea that the analytical threshold concept provides a measure of generosity and fairness toward the suspect. The wrong-headedness is the assumption that hiding subthreshold evidence benefits the suspect. It benefits *some* suspects, but those benefiting from hiding sub-threshold data tend to be the guilty ones. The simple reason is that the more information that is exposed, the better the chance that truth – innocence of the innocent or guilt of the guilty – comes to light. Truth is fodder for the innocent. Sub-threshold information is imperfect, but can usually be recognized as valid evidence when it is. To recap – the analytical threshold barrier tends to be beneficial for those suspects that we don't care much about, and tends to be harmful against the good guys.

So, there is a problem. Unfortunately fixing the Analytical Threshold concept it isn't so easy. It has a reason. But it isn't a very good remedy, so it needs to be improved.

B.2. The misunderstanding that sub-threshold information is unreliable

The common idea that sub-threshold data is unreliable, is also wrong. The idea makes sense for visible data, as a visible mark might represent an artifact, i.e. garbage, rather than an allele. But *lack* of any signal at an allele position – i.e. a void, ("Baskerville data", following mystery writer Sir Conan Doyle) is also evidence, and moreover it is reliable evidence. There is no garbage void, only reliable void. A void is evidence for example when it shows the absence of some suspect allele.

Results and conclusions

Analysis of a shell casing DNA mixture was presented in a current criminal case. The analysis, using a dropout-capable mixture program as recommended, calculated a hefty likelihood ratio of 825 connecting the defendant to the casing. But consideration of the sub-threshold pattern of the data led to the new insights and a likelihood ratio favoring the defendant.

The issue of Baskerville data arises with all mixture software that purports to deal with dropout and that filters the data with a threshold. That includes virtually all of the available continuous and semi-continuous model programs.

An occasionally stated dictum is the wisdom that while sub-threshold information should not be used as inculpatory evidence it can be used to exclude. How that might be done – void can be exculpatory – was not explained until now.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 61-68
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Individual Privacy and Reporting of Biological Sex

Vanessa Lynch[1], Laura Jane Heathfield[2], Bruce Budowle[3]

[1]*University of Stellenbosch, Cape Town, South Africa*
[2]*University of Cape Town, Cape Town, South Africa*
[3]*University of Helsinki, Helsinki, Finland*

** Corresponding author: Vanessa Lynch, University of Stellenbosch, vanessa@dnaforafrica.com, +27 83 2810509*

Abstract

Human identification by forensic DNA profiling primarily relies on the analysis of short tandem repeat markers (STRs) and Amelogenin or other sex-determining markers. Amelogenin serves as a phenotypic marker for biological sex, is typically considered innocuous, and the biological sex results are routinely reported and/or uploaded to national DNA databases. However, biological sex does not necessarily align with gender identity. Biological sex is genetically determined and assigned at birth based on anatomical features, whereas gender identity is an individual expression that may change over time and may not correspond with biological sex. Revealing differences between biological sex and gender identity may impinge on individual privacy. Furthermore, genetic anomalies related to the presentation of biological sex can occur, and the consequences of revealing such anomalies may have implications for the individuals involved. Disclosing biological sex in a report does not take into account ramifications for persons or their families with genetic anomalies related to sex chromosomes (which may or may not be known to the individual), transgender transformations (which may or may not have been disclosed by the individual), or gender-identity expressions that do not correspond with biological sex. While knowledge of biological sex may be important for operational forensic DNA laboratories and critical in certain cases, it is often not relevant to criminal investigations, courtroom deliberations, or public disclosure. It is incumbent upon us all to understand that the dissemination of biological sex data in the public domain, where disclosure is unnecessary, may impact individual privacy. In light of the current understanding and growing impact of gender identity, it is recommended that, with all due speed, (1) 'biological sex' and 'gender' be recognized as separate concepts, not to be used synonymously nor interchangeably; (2) definitions of a DNA profile be reviewed, and more clarity added; and (3) policies and protocols be developed to restrict such information from reports and court proceedings (i.e., public arenas) when not relevant.

Keywords
Gender; biological sex; privacy; Amelogenin; forensic DNA profiling.

Introduction

An important genetic marker, often not considered for individualisation, is the Amelogenin locus. Since the 1990s [1], Amelogenin has been used to determine the biological sex of a donors of crime scene samples and human remains, the proportions of female and male components in mixtures, and in kinship analyses. Biological sex may seem non-intrusive from a privacy perspective; however, it can contain personal information that could impact the health and social privacy of individuals. Where gender identity can differ from biological sex, the term "sex" refers to the biological sex and anatomical differences typically assigned to an individual at birth. In contrast, gender, or gender identity, is defined as "each person's deeply felt, internal and individual experience of gender, which may or may not correspond to the person's physiology or designated sex at birth" [2].

Excluding biological sex from the list of sensitive markers could achieve an appropriate balance between the intrusion of an individual's rights to privacy, human dignity, and equality, and the need to protect society against violent crimes. However, disclosing biological sex in the public domain may have ramifications for individuals, and their families, whose gender identity expressions do not align with their biological sex, transgender transformations, or individuals with genetic anomalies related to their sex chromosomes.

Biological sex marker results (such as Amelogenin, sex chromosome polymorphic loci, and targets for quantitative PCR assays) may be critical in some cases and thus should be part of the testing regimen. Moreover, one cannot know a priori when private information will be discovered with these markers or even be aware that there is a privacy concern. Therefore, striking a balance based on what to disclose and developing policies that uphold the values that seek to 'improve the quality of life of all citizens and free the potential of each person' [3] should be considered [4].

Distinction between biological sex and gender

Biological sex and gender are not synonymous, and the two concepts are not interchangeable. The important differences can be explained by considering three distinct elements related to an individual:

(i) Biological sex is determined at birth and does not change.
(ii) Gender identity is how one identifies oneself; it is fluid and may change throughout life.
(iii) Sexual orientation refers to whom a person is attracted to or loves; it is fluid and may change throughout life.

An individual's gender identity cannot be determined from DNA profiling data. Despite the distinct differences between the two concepts, they are frequently confused in scientific literature, legislation, academia, and general conversation. Conflating these two concepts makes notions of aberration from some sort of 'societal norm' relevant to many individuals in ways that can cause unwarranted and unnecessary distress or discomfort [5].

The biolegal process

Lynch and McNally [6] described the dynamic and symbiotic interaction between biotechnology and the laws required to legislate and govern it as 'biolegality.' Rights, identities, and the credibility of criminal evidence have been redefined in part due to this interaction. Building on the concept of biolegality, Wienroth et al. [7] further categorized the biolegal process prevalent in forensic DNA profiling into four "waves," ranging from the integrity of DNA evidence to the expansion of DNA databases for investigative/intelligence uses, and to advanced technologies such as massively parallel sequencing and RAPID DNA.

The addition of a "fifth" wave in the biolegal process

We introduce a 'fifth wave' to represent the continuation of advances and developments in science, highlighting the effects those advancements have on law as well as societal perspectives—i.e., a bio-sociolegal process (see Weinroth et al [7]. The

fifth wave is primarily concerned with the intersection of social factors and scientific evidence, and the way in which forensic molecular evidence is interpreted, stored, and displayed in reports, all of which impact privacy, equality, and human dignity. One focus area is the impact of gender theory conceptualizations and their consequences regarding genetic informational privacy associated with biological sex markers and their potential effects on an individual's privacy.

Genetic privacy

There are four types of genetic privacy pertaining to DNA: physical or bodily privacy, familial or relational privacy, special or locational privacy, and informational privacy [8]. Informational (genetic) privacy and human dignity have been discussed substantially and globally regarding the information contained in a forensic DNA profile. However, the notion that biological sex may impact privacy has not been raised or examined in a similar context.

Both gender identity and a person's genetic makeup are fundamental aspects of the self. If a person has not disclosed their biological sex or is unaware of genetic anomalies associated with expression of biological sex, disclosure of that information may interfere with the person's right to privacy, cause mental anguish or distress, or lead to discrimination. A core value of privacy lies in the right to choose what and how much personal information an individual shares with others. The concepts of personality, intimacy, and identity, in the context of a person's chosen gender identity and the desire to keep certain information about oneself private, are important components of the extended right to biological sex privacy. Therefore, this information should be protected [**9**]. While important for all of society, forensic scientists can be proactive and contribute through policies and practices to protect privacy that can be impacted by disclosing biological sex, especially in circumstances where biological sex is not germane to deliberations.

Biological sex privacy in the context of gender identity

A person's identity may diverge at the intersection of gender identity and biological sex. Gender identity may be openly expressed, while biological sex may remain private. Gender-diverse individuals undeniably suffer prejudice, discrimination, and even violence. In a report by the Human Rights Council (HRC) [**10**], the impor-

tance of not just *acknowledging* the rights of gender-diverse individuals, but also creating policies and laws that genuinely *protect* them, is highlighted [**10**].

The following cases serve as examples to raise awareness of instances where inadvertent private information regarding biological sex may be derived from DNA analyses and when such information should be considered private. (The first two cases have been anonymized and modified for privacy reasons.)

Case I: A False Lead from a Null Y Allele

In a murder investigation, a suspect's DNA profile was used to assist in determining the potential source of biological evidence from the crime scene. The Amelogenin marker revealed only an X allele, suggesting a possible, but not absolute, biological female identity. The Amelogenin marker did not alter or contribute to the association of the suspect, and the inferred biological sex evidence was not used for statistical calculations. Several explanations could account for this result: the suspect could be a biologically male individual with a null Y allele, a transgender person, or a male with a rare genetic anomaly such as De La Chapelle Syndrome [**11**]. The disclosure of this information in a public forum could unknowingly prejudice the defendant and affect court proceedings. This unnecessary exposure to sensitive personal information highlights the ethical dilemma of reporting biological sex in forensic contexts, as well as the reporting of potentially irrelevant information in a public setting (although the genetic information should be maintained in the laboratory case file).

Case II: Genetic Anomaly in a Paternity Case

During divorce proceedings, a court mandated a DNA paternity test for a child, which excluded the alleged father. The results also indicated that he was a 46,XX (SRY+) male, suggesting that he was infertile [11]. Although one could assert that his infertility supported the claim of non-paternity, this disclosure was unnecessary, as the standard aSTRs were sufficient to exclude him as the biological father. However, while not part of the purpose of the paternity test, the revelation of infertility raised questions about the paternity of his other children. This case warrants further discussion within the forensic genetics community regarding proper policies on the disclosure of sex chromosome anomalies in such cases, particularly

when this information may not be necessary for rendering an effective interpretation and may, more importantly, impact other family members who were not part of the initial testing question.

Case III: Importance of Biological Sex in Forensic Identification

Initial assessments of unidentified human remains found on a beach suggested that the remains were biologically female. In contrast, DNA analysis confirmed that both sets of remains were biologically male, refocusing the police investigation. In this case, determining biological sex was critical for the investigation.

Conclusion

Biological sex marker results are informative in a small portion of cases and are often helpful in a supporting role in the interpretation of DNA results within the laboratory. In most situations, biological sex does not need to be revealed in reports or in the presentation of DNA evidence in court, and it would rarely, if ever, be needed for exigent circumstances related to the safety and security of society. However, biological sex is routinely shown in forensic reports and presented in public forums. The disclosure of the biological sex of a suspect, victim, and/or other persons associated with the investigation can have potentially adverse consequences for the individuals concerned, as well as their family members. The discrimination and biases faced by such individuals are largely influenced by the sharing of an individual's biological sex.

Because of genetic anomalies and the right to choose one's gender identity, along with the consequences of disclosure, privacy needs to be protected. Taking a leadership role in safeguarding privacy should be part of the ethos of forensic sciences, especially when dealing with genetic data. The responsibility for setting policies and practices can be undertaken by the forensic science community, particularly when the disclosure of biological sex has little or no impact on a case. Laboratories could take the lead by collaborating with investigators and the judicial community to develop policies and practices that address the needs of litigation while safeguarding and/or redacting sensitive information.

It is imperative to consider biological sex markers as sensitive genetic data that may impact privacy, as the information in particular cases, while not relevant,

may contribute to the potential prejudice already experienced by vulnerable minorities, including the gender-diverse community. The forensic science community should, with all due speed, consider establishing practices that are commensurate with current understandings of the differences between biological sex and gender identity.

Declaration of competing interests:

None

References

[1] K.M. Sullivan, A. Mannucci, C.P. Kimpton, P. Gill, A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene Amelogenin, Biotechniques 15(4) (1993) 636-641.

[2] World Health Organization "Gender and Health" (2023), Available at: https://www.who.int/health-topics/gender#tab=tab_1

[3] Preamble to the South African Constitution, Available at: https://www.justice.gov.za/legislation/constitution/preamble.html#:~:text=We%2C%20the%20people%20of%20South%20Africa%2C%20Recognise%20the,who%20live%20in%20it%2C%20united%20in%20our%20diversity.

[4] L.J. Heathfield. Policy required for entry of DNA profiles onto the National Forensic DNA Database of South Africa. South African Journal of Science, 110(7/8) (2014) 1-3.

[5] R. Mcdermott, P. Hatemi, Distinguishing sex and gender, Political Science and Politics 44 (1) (2011) 89-92.

[6] M. Lynch, R. McNally, Forensic DNA databases: The co-production of law and surveillance technologies, Handbook of Genetics and Society. Mapping the New Genomic Era, P. Atkinson, P. Glasner and M. Lock (eds), Routledge, London (2009) pg. 284.

[7] M. Wienroth, N. Morling, R. Williams, Technological innovations in forensic genetics: Social, legal and ethical aspects, Recent Advances in DNA and Gene Sequences 8(2) (2014) 98-103.

[8] S. Krimsky, T. Simoncelli, Genetic justice: DNA data banks, criminal investigations, and civil liberties, Columbia University Press, New York, (2011) pp. 228-232.

[9] De Cew "Privacy" in *The Stanford Encyclopedia of Philosophy*.

[10] UNGA Report of the Independent Expert on protection against violence and discrimination based on sexual orientation and gender identity, Victor Madrigal-Borloz, UN Doc A/HRC/47/27 para 36, 41. The HRC replaced the former United Nations Commission on Human Rights. See, United Nations Human Rights Council "Welcome to the Human Rights Council" (2020) UNHRC <https://www.ohchr.org/en/hrbodies/hrc/pages/aboutcouncil.aspx> (accessed 08-11-2021).

[11] A. de la Chapelle, Analytic review: nature and origin of males with XX sex chromosomes, Am. J. Hum. Genet. 24(1) (1972) 71–105.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 69-76
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Dealing with incidental findings in Forensic Genetics

Martin Zieger[1] *

[1] *Institute of Forensic Medicine, Forensic Molecular Biology Department, University of Bern, Murtenstrasse 26, 3008 Bern, Switzerland*

** Corresponding author at: martin.zieger@irm.unibe.ch*

Abstract

Following the impressive development of sequencing techniques in the last decade, the analytical scope of forensic genetics has significantly broadened with, among others, the validation and legal approval of Forensic DNA Phenotyping, mitochondrial whole genome analysis and Investigative Genetic Genealogy based on large-scale SNP array data. It can be assumed that this development has not yet reached its end. It is evident that privacy intrusion becomes more serious and that data protection issues become more important when more genetic data is analyzed. Particularly with the paradigm shift from the analysis of non-coding to coding regions of DNA, we can foresee that incidental findings with potential health implications for the concerned individuals will become more frequent in the future, even though efforts are made in forensic genetics to avoid such information to arise. The treatment of incidental findings seems to have been largely neglected in forensic genetics, whereas human geneticists in the medical field are well aware of the problem and have developed guidelines for the correct handling of incidental findings. This contribution gives a brief overview over international regulations and policies concerning the treatment of incidental findings and gives some practical examples for incidental findings in forensic genetics. The article formulates a call for the development of guidelines in the field, to improve data fairness for convicted offenders, suspects, victims and for other individuals from whom genetic data may be generated for forensic purposes.

Keywords

Incidental findings, secondary findings, fundamental rights, policy, guidelines, forensic genetics.

Introduction

The *right to know* as well as the *right not to know* about the results of a genetic examination are well-established ethical principles, as stated by article 5 of the UNESCO Declaration on the Human Genome and Human Rights in 1997 [1]. In 2003, the UNESCO Declaration on Human Genetic Data in article 13 more specifically established the principle that no one should be denied access to their genetic data, unless restricted by domestic law, for compelling reasons [2]. Those principles are mentioned in several other international legal documents, such as for example in article 10(2) of the Council of Europe's Oviedo Convention on Human Rights and Biomedicine of 1997 [3] and the related *right of access* is also an integral part of numerous data protection legislations, such as the European Union's General Data Protection Regulation, where it is explicitly mentioned in article 15 [4]. However, implementation of those rights on the domestic level is extremely heterogeneous and a plethora of different policies exists [5].

Concerning international policies, it seems that a consensus emerges towards the communication of incidental findings from genomics research as well as from clinical genomics, with a continuously growing list of genetic variants that are recommended for reporting [6, 7]. The influential Global Alliance for Genomics and Health (GA4GH) encourages in their 2021 Policy on Clinically Actionable Genomic Research Results the return of results in the research domain, respecting participant's choices [8, 9]. In clinical genomics, the World Medical Association's Declaration of Reykjavik of 2019 encourages the return of secondary / incidental findings if they are reliable, concern a severe health condition and if medical action is possible. In principle, patient's choice on the return of results should be respected. However, this choice might even be overridden if the findings have severe implications for family members [10].

Compared to whole genome sequencing in genomics research or the application of variant panels in clinical genetics, genetic analysis for forensic purposes is less prone to generate medically actionable incidental findings. However, it is a fallacy to assume that we are not affected at all in our field. In the following, I will present three case examples from different contexts that could be encountered in most forensic genetics labs.

Results and Discussion

The examples presented in this section are selected to illustrate that incidental findings can arise in various situations in a forensic genetics lab. However, the following section requires an important disclaimer in advance: The recommendations given for every case are not meant to be legal advice. They are just suggestions for a minimal standard for being in line with the mentioned international documents. Since legislation is extremely variable on the domestic level, it is not possible to give universal legal recommendations. For example, the Swiss law prohibits the communication of all unintended findings for paternity cases, as well as in the forensic domain [11, 12].

a. Case 1: Method validation

Case circumstances:

Imagine you want to establish a new protocol in your lab to implement whole mtDNA sequencing for the forensic analysis of degraded human remains. You approach your staff for buccal swab samples for the validation study. Among the participants, you have a 25-year-old female collaborator who gives a sample.

Findings:

In the sample of your healthy 25-year-old colleague, you discover an m.3243A>G heteroplasmy on the mitogenome, a relatively frequent variant with an extremely variable degree of manifestation. The pathogenic nucleotide variant impairs the functioning of the mitochondrial tRNA for leucin [13]. Obviously, your colleague is not affected by this potentially pathogenic variant. However, she might want to have children in the next years and you know that pathogenicity of the m.3243A>G heteroplasmy is bound to the heteroplasmy level that could change in the gametes, leading to serious health impairment in their offspring [14-16]. Prenatal screening or preimplantation diagnostic might be warranted and even a mitochondrial replacement therapy could be an option.

Consequences for the case:

Even though the finding of the m.3243A>G heteroplasmy appears to have no implications for the health of your collaborator, it can have large implications for her personal live, by potentially affecting her family planning. Communication of the finding might prevent harm from your collaborator and her offspring. The finding would be actionable. The information about the possibility that such a finding could arise should be part of the informed consent documentation. It might be difficult in the present constellation to respect the right not to know. Since the

concerned individual is working in the same place where the genetic results are generated, it might be inevitable that she will become aware of them. In general, caution is warranted when asking collaborators for samples, since free consent could be impaired by felt obligations to participate towards the superior.

b. Case 2: Paternity testing

Case circumstances:

A couple presents at your lab to perform a paternity test for their male baby.

Findings:

In the sample of the baby, you discover a peak imbalance at the Amelogenin locus, highly indicative of a Klinefelter syndrome [17, 18]. Klinefelter is relatively common, with an estimated frequency of 1 in 500 to 1 in 1,000 men. However, it is also estimated that most of the affected individuals never get it diagnosed. Diagnosis is most often made upon consultation with a reproductive medical department, due to the associated infertility caused by azoospermia [19]. Klinefelter syndrome can affect the health of the concerned individuals in various ways and treatments such as testosterone substitution to improve health are available. Infertility might be overcome through testicular sperm extraction and cryopreservation of extracted sperm cells [20, 21].

Consequences for the case:

The paternity test revealed a clinically actionable finding. Before conducting the test, you should inform the parents that such a finding might occur and ask them whether they wish to be informed of it. Most likely, your forensic genetics lab has not the expertise for a detailed genetic counselling, with regard to clinically relevant findings and, even though highly indicative, your test result is not valid for diagnosing Klinefelter. You should therefore refer the parents to a clinical grade genetic testing lab for confirmation and counselling in case of an incidental finding. A good alternative would be to avoid the analysis of the Amelogenin locus, by using a multiplex PCR kit without sex marker, as proposed previously [18].

c. Case 3: Forensic DNA phenotyping

Case circumstances:

A woman is found stabbed in the living room of her apartment. A bloodstain is found in the bathroom from which a male DNA profile is established. Since the profile does not lead to a hit in the national DNA database for convicted offenders, it is subjected to a phenotyping analysis and age prediction based on methylation patterns [22].

Findings:

The age of the individual who left the bloodstain is estimated at 42 years. Through further police investigation and subsequent conventional DNA profiling, it is found that the bloodstain had been caused by a plumber repairing the shower in the week before the murder. The plumber is aged 24. A large upward deviation of the age predicted from the methylation patterns of the sample from the actual chronological age of the concerned individual (the plumber) could be caused by a severe disease [23, 24].

Consequences for the case:

The finding of potential premature biological ageing could be of great significance to the plumber, since this could be caused by a severe actionable disease, such as leukemia [23]. However, it is of unclear clinical significance. It might nevertheless be fair to alert the plumber of this finding in a general way and to recommend a medical check-up with a physician. The result of the forensic age prediction could also be communicated directly to a physician, designated by the concerned individual. Most likely, this information would have to be communicated by the contracting law enforcement agency, since in most cases the forensic genetics lab will not have sufficient information on the concerned individual. Whether action should be taken if unintended findings arise, could be asked upon the sampling of the reference buccal swab that is needed to eventually attribute the bloodstain, and therefore also the results of the age prediction, to the plumber. However, data fairness through access to personal genetic data and the realization of the right to know for forensic samples are challenging and the issue has been largely underexplored to date.

Conclusion

This contribution aimed to demonstrate through a few examples that incidental, actionable findings can arise in forensic genetics. The examples given are by no means exhaustive. In our labs, we generate genetic data in various different contexts, all requiring a slightly different handling. The list includes:

a. Research
b. Paternity / kinship testing
c. Staff samples for elimination databases
d. Reference samples from victims, volunteers, etc.
e. Samples from arrestees / suspects and convicted offenders

f. Deceased individuals
g. Data from crime scene samples attributed to individuals

Some countries may have detailed regulation on how to handle the genetic data generated in all these different contexts. However, most likely legal guidance is largely absent and a policy document could be warranted and helpful to the community. The present contribution is therefore also a call for an open debate.

Conflict of interest statement

The author has no conflict of interest to declare.

References

[1] UNESCO, Universal Declaration on the Human Genome and Human Rights, 1997.
[2] UNESCO, International Declaration on Human Genetic Data, 2003.
[3] Council of Europe, Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine, Oviedo, 1997.
[4] Regulation (EU) 2016/679 of the European Parliament and of the Council: On the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (2016), OJ L119/1, (GDPR), Art. 9(2)(j).
[5] A. Thorogood, G. Dalpé, B.M. Knoppers, Return of individual genomic research results: are laws and policies keeping step?, European Journal of Human Genetics 27(4) (2019) 535-546.
[6] S.M. Wolf, R.C. Green, Return of Results in Genomic Research Using Large-Scale or Whole Genome Sequencing: Toward a New Normal, Annual Review of Genomics and Human Genetics 24(Volume 24, 2023) (2023) 393-414.
[7] D.T. Miller, K. Lee, N.S. Abul-Husn, L.M. Amendola, K. Brothers, W.K. Chung, M.H. Gollob, A.S. Gordon, S.M. Harrison, R.E. Hershberger, T.E. Klein, C.S. Richards, D.R. Stewart, C.L. Martin, ACMG SF v3.2 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG), Genetics In Medicine 25(8) (2023) 100866.

[8] Global Alliance for Genomics and Health: 2021 Policy on Clinically Actionable Genomic Research Results, 2021.

[9] A.C.F. Lewis, B.M. Knoppers, R.C. Green, An international policy on returning genomic research results, Genome Med 13(1) (2021) 115.

[10] World Medical Association (WMA) Declaration of Reykjavik – Ethical Considerations Regarding the Use of Genetics in Health Care, 2019.

[11] M. Zieger, Abstammungsuntersuchungen nach dem revidierten GUMG, Jusletter (2019).

[12] M. Zieger, Forensische DNA-Analyse: So viel wie nötig, so wenig wie möglich?, Jusletter (2020).

[13] H.R. Elliott, D.C. Samuels, J.A. Eden, C.L. Relton, P.F. Chinnery, Pathogenic Mitochondrial DNA Mutations Are Common in the General Population, The American Journal of Human Genetics 83(2) (2008) 254-260.

[14] D. Li, C. Liang, T. Zhang, J.L. Marley, W. Zou, M. Lian, D. Ji, Pathogenic mitochondrial DNA 3243A>G mutation: From genetics to phenotype, Front Genet 13 (2022) 951185.

[15] J.B. Stewart, P.F. Chinnery, The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease, Nature Reviews Genetics 16(9) (2015) 530-542.

[16] J.P. Grady, S.J. Pickett, Y.S. Ng, C.L. Alston, E.L. Blakely, S.A. Hardy, C.L. Feeney, A.A. Bright, A.M. Schaefer, G.S. Gorman, R.J. McNally, R.W. Taylor, D.M. Turnbull, R. McFarland, mtDNA heteroplasmy level and copy number indicate disease burden in m.3243A>G mitochondrial disease, EMBO Molecular Medicine 10(6) (2018) e8262.

[17] A. Spitzer, L. Elkayam Sapir, M. Amiel, What is she doing here? Klinefelter syndrome in forensic casework, Science & Justice 61(4) (2021) 443-448.

[18] A. Gabriele, E. Chierto, S. Gino, S. Inturri, S. Aneli, C. Robino, Privacy and ethical challenges of the Amelogenin sex test in forensic paternity/kinship analysis: Insights from a 13-year case history, Forensic Science International: Synergy 7 (2023) 100440.

[19] K.A. Groth, A. Skakkebæk, C. Høst, C.H. Gravholt, A. Bojesen, Clinical review: Klinefelter syndrome--a clinical update, J Clin Endocrinol Metab 98(1) (2013) 20-30.

[20] M. Zitzmann, L. Aksglaede, G. Corona, A.M. Isidori, A. Juul, G. T'Sjoen, S. Kliesch, K. D'Hauwers, J. Toppari, J. Słowikowska-Hilczer, F. Tüttelmann, A. Ferlin, European academy of andrology guidelines on Klinefelter Syndrome Endorsing Organization: European Society of Endocrinology, Andrology 9(1) (2021) 145-167.

[21] G. Butler, U. Srirangalingam, J. Faithfull, P. Sangster, S. Senniappan, R. Mitchell, Klinefelter syndrome: going beyond the diagnosis, Archives of Disease in Childhood 108(3) (2023) 166-171.

[22] M. Kayser, W. Branicki, W. Parson, C. Phillips, Recent advances in Forensic DNA Phenotyping of appearance, ancestry and age, Forensic Science International: Genetics 65 (2023) 102870.

[23] M. Spolnicka, R. Zbiec-Piekarska, M. Karp, M.M. Machnicki, P. Wlasiuk, Z. Makowska, A. Pieta, T. Gambin, P. Gasperowicz, W. Branicki, K. Giannopoulos, T. Stoklosa, R. Ploski, DNA methylation signature in blood does not predict calendar age in patients with chronic lymphocytic leukemia but may alert to the presence of disease, Forensic Sci Int Genet 34 (2018) e15-e17.

[24] M. Spolnicka, E. Pospiech, B. Peplonska, R. Zbiec-Piekarska, Z. Makowska, A. Pieta, J. Karlowska-Pik, B. Ziemkiewicz, M. Wezyk, P. Gasperowicz, T. Bednarczuk, M. Barcikowska, C. Zekanowski, R. Ploski, W. Branicki, DNA methylation in ELOVL2 and C1orf132 correctly predicted chronological age of individuals from three disease groups, International Journal of Legal Medicine 132(1) (2018) 1-11.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 77-84
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Evaluation of DNA from teeth subjected to various extreme ante-mortem degradation factors: preliminary study

Ana Maria Salazar[a, b], Ariel Castro[c], Patricia Huerta[d], Anna Barbaro[e, f, g*]

[a] *University of Concepción, Department of Pathology and Diagnosis, Faculty of Dentistry, Concepción, Chile.*

[b] *University of Alcalá, PhD Student in Forensic Sciences.*

[c] *University of Concepción, Department of Biochemistry and Molecular Biology, Faculty of Biological Sciences, Concepción, Chile.*

[d] *University of Concepción, Department of Public Health, Faculty of Medicine, Concepción, Chile.*

[e] *University of Alcalá, Department of Analytical Chemistry, Physical Chemistry, and Chemical Engineering, Madrid-Barcelona Road km 33.6, 28871 Alcalá de Henares, Madrid, Spain.*

[f] *Studio Indagini Mediche E Forensi (SIMEF), via Nicolò da Reggio 4, 89128 Reggio Calabria, Italy.*

[g] *University of Alcalá, University Research Institute in Police Sciences, Libreros 27, 28801 Alcalá de Henares, Madrid, Spain.*

** Corresponding author: a.barbaro@uah.es*

Abstract

In complex situations that require the analysis of highly degraded tissues, DNA is often the only viable option for identification, with teeth being a preferred source due to their durability. Ante-mortem factors like pathologies and tooth type can affect DNA recovery. While molars and premolars in ideal conditions are the most suitable, this study explores the potential for DNA recovery in "non-ideal" teeth with conditions such as caries, periodontitis, and endodontic treatments. **Material and Methods:** Ten samples of various tooth types and conditions were collected from the University of Concepción's Oral and Maxillofacial Surgery clinic. DNA was extracted, quantified using spectrophotometry, and evaluated for quality and quantity, followed by statistical analysis. **Results and Conclusion:** DNA concentration and quality showed no significant differences regarding tooth type, pathologies, or treatments. Premolars yielded the highest DNA, while canines produced the least and lowest quality. Teeth with caries provided less DNA, but those with periodontitis and endodontic treatments showed higher values. A statistical simulation aligned with results from other studies. In conclusion, while preliminary, the findings suggest trends that warrant further research with larger sample sizes and real-time PCR for more precise results.

Keywords

DNA, Teeth, Degradation, Ante-mortem factors

1. Introduction

DNA analysis is crucial in tragic events involving highly degraded or fragmented tissues, often serving as the only identification method. Teeth and bones are the preferred sources due to their resistance to degradation [1]. Teeth, in particular, provide exceptional durability thanks to their composition, location, and morphology, protecting tissues like pulp, dentin, and cementum even in extreme environments [2-8].

Factors influencing DNA recovery from teeth can be categorized into ante-mortem and post-mortem factors. Ante-mortem factors include tooth type, which affects the amount of cellular tissue available for DNA extraction, and pathological conditions such as caries, infections, or mechanical damage [7, 9-12]. Molars and premolars without pathologies are most suitable for DNA recovery [7, 13, 14].

This research investigates whether DNA can still be effectively recovered from teeth with "non-ideal" ante-mortem conditions.

2. Material studied, methods, techniques

2.1. Samples and data.

Ten tooth samples were collected from patients undergoing therapeutic extractions at the Maxillofacial Surgery clinic, Faculty of Dentistry, University of Concepción, Chile. Ethical protocols from the Scientific Ethical Committee of the Concepción Health Service were followed. Donor age, sex, tooth type, and associated pathologies (non-mutually exclusive) were documented. **Figure 1** outlines the teeth's characteristics and conditions at extraction.

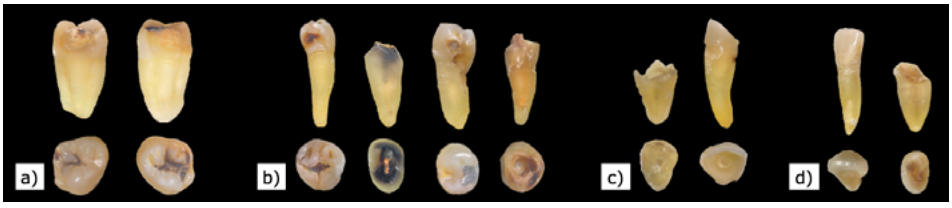


Figure 1: Characteristics of teeth and their condition at the time of extraction (N10)
a) Molars, b) Premolars, c) 2 Canines, d) 2 Incisors.

2.2. Preparation Procedures, DNA Extraction, and DNA Quantification.

All samples were decontaminated with 3% sodium hypochlorite for up to 1 minute, followed by sterile water washes. Decalcification was performed with EDTA for 15 days to remove mineralized substance. The tooth crowns were removed, leaving only the roots, which were ground into fine powder after immersion in liquid nitrogen using a mortar.

DNA was extracted from the powdered samples using the QIAamp DNA Investigator Kit (Qiagen), following the manufacturer's protocol [15]. DNA quality (260/280 ratio) and quantity (ng/µL) were assessed by spectrophotometry.

2.3.Statistical Analysis

Statistical analysis was then performed using the IBM Statistical Package for the Social Sciences (SPSS) V2 statistical program.

Results

We observed that several samples were destroyed by caries and in a root remnant state (**Fig. 1**); even in those conditions, it was possible to extract DNA.

The different statistical analyses indicate that there are no significant differences in the concentration and quality of DNA in relation to the type of tooth, pathologies, or treatments (T test, ANOVA, **tables 1 and 2**). However, less DNA was obtained from teeth with caries, while in cases of periodontal disease and endodontics treatments, the values were higher. The teeth from which the most DNA could be extracted were premolars. The pieces from which the least DNA was extracted were canines. However, the quality of DNA was better in canines (**table 1**).

A statistical simulation was performed with 10,000 resamplings using the BCa method to calculate confidence intervals, indicating that the confidence intervals would be similar (**tables 1 and 2**).

Table 1: Statistical analysis according to type of tooth

Independent variables (Type of tooth) p=0.456 (ANOVA)	N	Real samples (N10)								Statistical simulation (N10.000)[a]	
		DNA concentration (ng/uL)				Ratio 260/280				DNA Concentration (ng/uL)	Ratio 260/280
		Media	SD	Min	Max	Media	SD	Min	Max	CI 95%[b] (lower/upper)	CI 95%[b] (lower/upper)
Molar	2	10.8	2.4	9.1	12.5	1.505	0.049	1.470	1.540	(-)10.801/32.401	1.060/1.949
Premolar	4	20.6	13.5	5.7	33.3	1.433	0.554	0.620	1.810	(-)0.845/42.095	0.550/2.314
Canine	2	7.2	3.0	5.0	9.3	1.655	0.191	1.520	1.790	(-)20.168/34.468	(-)0.060/3.370
Incisor	2	10.9	7.6	5.5	16.3	1.430	0.212	1.280	1.580	(-)57.714/79.514	(-)0.475/3.335

a: Bootstrap N10000, b: CI BCa

Table 2: Statistical analysis according to dental pathology and treatment

Independent variables (Pathologies)	N	Real samples (N10)								Statistical simulation (N10.000)[a]	
		DNA concentration (ng/uL)				Ratio 260/280				DNA Concentration (ng/uL)	Ratio 260/280
		Media	SD	Min	Max	Media	SD	Min	Max	CI 95%[b] (lower/upper)	CI 95%[b] (lower/upper)
Caries p=0.158 (T test)	Yes (8)	11.7	9.3	5.0	33.3	1.446	0.375	0.62	1.81	7.6/16.333	1.173/1.668
	No (2)	23.5	10.2	16.3	30.7	1.670	0.127	1.58	1.76	16.3/30.7	1.580/1.760
Periodontal disease p=0.131 (T test)	Yes (6)	16.7	12.6	5.0	33.3	1.623	0.206	1.28	1.81	10.1/20.971	1.520/1.710
	No (4)	10.0	3.3	5.7	12.8	1.293	0.450	0.62	1.54	7.74/12.6	0.974/1.540
Endodontic treatment p=0.23 (T test)	Yes (2)	23.1	14.5	12.8	33.3	1.215	0.841	0.62	1.81	12.8/33.3	0.620/1.810
	No (8)	11.8	8.6	5.0	30.7	1.560	0.161	1.28	1.79	7.386/16.334	1.485/1.630

a: Bootstrap N10000, b: CI BCa

Discussion

The literature suggests that premolars and molars without pathologies or dental treatments are ideal for successful DNA analysis, as most studies focus on these samples [7, 9, 10, 13, 14, 16]. However, when only "non-ideal" samples are available, success rates can vary significantly [7, 10, 16, 17]. This preliminary study analyzed teeth with "non-ideal" ante-mortem factors, including various tooth types with pathologies (caries, periodontal disease) and endodontic treatments, and compared these with DNA yield.

The study found that DNA could be extracted from all samples, regardless of conditions. Most published studies [7, 9, 10,13, 14, 16], indicate that the main candidates for the selection of dental samples would be molars, followed by premolars, due to their morphology and the amount of cellular tissue available for DNA extraction (cementum and pulp). On the other hand, some authors [9, 16, 18], indicate that there are no significant differences in DNA extraction based on the type of dental piece. In our study, premolars yielded the most DNA, canines yielded the least, aligning with prior studies [7, 9, 10, 13, 14, 16]. However, canines showed better DNA quality, possibly due to associated pathology or treatment, as observed in similar research by Heathfield et al. [16].

Pathologies had varying effects on DNA extraction. Caries significantly hindered DNA yield, whereas teeth with endodontic treatments and periodontal disease produced good quality DNA. Higgins et al. [7] highlight that cementum remains a reliable DNA source in carious teeth, as it is unaffected by such pathologies. Similarly, endodontic treatments remove infected organic tissue, potentially preserving DNA quality.

For periodontal disease, Heathfield et al. [16] found increased DNA quantity and quality due to heightened blood flow and immune activity during inflammation. This study confirmed these findings, with teeth affected by periodontal disease yielding higher quality and quantity of DNA, similar to teeth with endodontic treatments. However, in advanced cases, cementum exposure could lead to cariogenic activity and the destruction of cementocytes [7].

In conclusion, DNA recovery is possible from non-ideal tooth samples, but pathologies must be carefully considered. Tooth selection for DNA extraction should be case-specific.

Conclusion

The complexity of DNA extraction from teeth arises from numerous interconnected variables, making comprehensive studies challenging. These factors contribute to a multifactorial degradation process. While this research had a limited sample size, future studies should include more samples and diverse methodologies, such as manual and automated DNA extraction and Real-Time PCR quantification. Despite its limitations, the preliminary results provide valuable insights and suggest trends for further investigation.

Conflict of interest statement

The authors declare no conflicts of interest.

References

[1] Budowle B, Bieber FR, Eisenberg AJ, Forensic aspects of mass disasters: strategic considerations for DNA-based human identification. Leg Med (Tokyo),(2005) Jul;7(4):230-43. doi.org/10.1016/j.legalmed.2005.01.001. Epub 2005 Mar 17. PMID: 15975517.

[2] Nanci A. 2003. Ten Cate's Oral Histology, Development, Structure and Function. 6a ed., Mosby, St. Louis, Missouri. Cap. 7: Enamel: Composition, Formation, and Structure. págs: 154, 176, 177.

[3] Sweet DJ, Sweet CH, DNA analysis of dental pulp to link incinerated remains of homicide victim to crime scene. J Forensic Sci, (1995) Mar;40(2):310-4. PMID: 7602298.

[4] Mörnstad H, Pfeiffer H, Yoon C, et al, Demonstration and semi-quantification of mtDNA from human dentine and its relation to age. Int J Legal Med, (1999);112(2):98-100. doi.org/10.1007/s004140050209. PMID: 10048666.

[5] Shiroma CY, Fielding CG, Lewis JA Jr, et al, A minimally destructive technique for sampling dentin powder for mitochondrial DNA testing, J Forensic Sci, (2004) Jul;49(4):791-5. PMID: 15317196.

[6] Vavpotic M, Turk T, Martincic DS, et al, Characteristics of the number of odontoblasts in human dental pulp post-mortem, Forensic Sci Int., (2009) Dec 15;193(1-3):122-6. doi.org/10.1016/j.forsciint.2009.09.023. Epub (2009) Nov 4. PMID: 19892501.

[7] Higgins D, Austin JJ, Teeth as a source of DNA for forensic identification of human remains: a review. Sci Justice, (2013) Dec;53(4):433-41. doi.org/10.1016/j.scijus.2013.06.001. Epub 2013 Jul 2. PMID: 24188345.

[8] D. B. N, Joji JG, Tooth as a Source of DNA in Forensic or Investigative Genetics: An Overview. jfds [Internet], (2022) Jul. 25 [cited 2023 Jul. 21];12(3):186-9. Available from: https://www.jfds.org/index.php/jfds/article/view/617

[9] Mansour H, Krebs O, Pinnschmidt HO, et al, Factors affecting dental DNA in various real post-mortem conditions. Int J Legal Med, (2019) Nov;133(6):1751-1759. doi.org/10.1007/s00414-019-02151-9. Epub (2019) Sep 11. PMID: 31511983.

[10] Rubio L, Martinez LJ, Martinez E, et al, Study of short- and long-term storage of teeth and its influence on DNA. J Forensic Sci, (2009) Nov;54(6):1411-3. doi.org/10.1111/j.1556-4029. (2009).01159.x. Epub 2009 Oct 5. PMID: 19804532.

[11] Ricucci D, Siqueira JF Jr, Loghin S, et al, Pulp and apical tissue response to deep caries in immature teeth: A histologic and histobacteriologic study. J Dent.,(2017) Jan;56:19-32. doi.org/10.1016/j.jdent.2016.10.005. Epub(2016) Oct 12. PMID: 27744048.

[12] Carvalho TS, Colon P, Ganss C, et al, Consensus Report of the European Federation of Conservative Dentistry: Erosive tooth wear – diagnosis and management. Swiss Dent J, (2016);126(4):342-346. PMID: 27142130.

[13] INTERPOL Disaster Victim Identification Guide [Internet]; 2018 [Consulted 2023, Mayo 9] Available in: https://www.interpol.int/How-we-work/Forensics/Disaster-Victim-Identification-DVI

[14] Prinz M, Carracedo A, Mayr WR, et al, International Society for Forensic Genetics. DNA Commission of the International Society for Forensic Genetics (ISFG): recommendations regarding the role of forensic genetics for disaster victim identification (DVI). Forensic Sci Int Genet.,(2007) Mar;1(1):3-12. doi.org/ 10.1016/j.fsigen.2006.10.003. Epub (2006) Nov 28. PMID: 19083722.

[15] https://www.qiagen.com/es-us/products/human-id-and-forensics/investigator-solutions/qiaamp-dna-investigator-kit. Access: 16.08.2023

[16] Heathfield LJ, Haikney TE, Mole CG, et al, Forensic human identification: Investigation into tooth morphotype and DNA extraction methods from teeth. Sci Justice,(2021) Jul; 61(4):339-344. doi.org/10.1016/j.scijus.2021.05.005. Epub (2021) May 30. PMID: 34172122.

[17] Baeta M, Núñez C, Cardoso S, et al, Digging up the recent Spanish memory: genetic identification of human remains from mass graves of the Spanish Civil

War and posterior dictatorship. Forensic Sci Int Genet, (2015) Nov;19:272-279. doi.org/10.1016/j.fsigen.2015.09.001. Epub 2015 Sep 7. PMID: 26378723.

[18] Corte-Real A, Anjos MJ, Vieira DM, et al, The tooth for molecular analysis and identification: a forensic approach. J Forensic Odontostomatol, (2012) Jul 1;30(1):22-8. PMID: 23000808; PMCID: PMC5734848.

30th INTERNATIONAL SOCIETY FOR FORENSIC GENETICS CONGRESS
Universidade de Santiago de Compostela, 2025, pp. 85-91
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Insertion-Null markers: Efficacy in degraded bones analysis

Christian Haarkötter[1], María Saiz[1], Xiomara Gálvez[1], Diana C. Vinueza-Espinosa[1], María Isabel Medina-Lozano[1], Juan Carlos Álvarez[1]*, Jose Antonio Lorente[1]

[1] *Laboratory of Genetic Identification & Human Rights (LABIGEN-UGR), Department of Legal Medicine, Faculty of Medicine, University of Granada, Av. Investigación 11 – PTS – 18016, Granada, Spain.*

** Corresponding author at: juanca@ugr.es*

Abstract

Forensic DNA analysis of compromised skeletal remains is often hindered by DNA degradation, leading to incomplete or negative autosomal STR profiles. Alternative methods like mitochondrial DNA or SNP typing can be used, but they are typically labor-intensive and expensive. Insertion-null alleles (INNULs), which are short interspersed nuclear elements, have been proposed as effective tools for human identification in difficult samples due to their small amplicon size. A commercial kit, InnoTyper® 21, which includes 20 INNUL markers and amelogenin, has been developed for this purpose. This study [1]often resulting in partial or negative autosomal STRs profiles. To address this issue, alternative approaches such as mitochondrial DNA or SNPs typing may be employed; however, they are labour-intensive and costly. Insertion-null alleles (INNULs evaluates its effectiveness on degraded skeletal remains by comparing the number of detected alleles, RFU values, PHR, and reportable markers with those obtained using GlobalFiler™. The random match probability for each sample was calculated using Familias version 3 to assess the discriminatory power of each kit. InnoTyper® 21 consistently produced more alleles, higher RFU values, and more reportable loci. However, both profiles were generally equally informative. In conclusion, InnoTyper® 21 is a valuable addition for analyzing challenging samples, especially when routine methods yield poor or negative profiles.

Keywords

Forensic Genetics, Human remains, INNULs, Retrotransposable elements, STRs.

Introduction

Analyzing forensic DNA from degraded skeletal remains using STRs is challenging due to DNA fragmentation and molecular damage caused by decomposition and environmental factors [2]. Commercial STR kits, which typically produce amplicons between 100 and 500 base pairs (bp), may yield partial or negative profiles due to DNA degradation. To mitigate this, strategies like miniSTRs or mtDNA analysis are employed.

Insertion and deletion polymorphisms, which lie between STRs and SNPs, are advantageous as they are compatible with standard capillary electrophoresis workflows. Retrotransposable elements (REs) include long and short interspersed nuclear elements (LINEs and SINEs). Among SINEs, Alu sequences feature insertion and null alleles (INNULs) of varying lengths. A commercial kit, InnoTyper® 21, includes 20 INNUL markers and uses a three-primer strategy (a common forward primer for both alleles and specific primers for insertion and null alleles).

INNUL typing offers benefits such as small amplicon size, absence of stutter artifacts, and a low mutation rate. However, a notable drawback is the significant length difference between insertion and null alleles, leading to preferential amplification of the smaller allele. These features make INNULs useful for human identification, analyzing degraded samples, interpreting mixtures, and conducting population and biogeographical ancestry studies.

This research [1]often resulting in partial or negative autosomal STRs profiles. To address this issue, alternative approaches such as mitochondrial DNA or SNPs typing may be employed; however, they are labour-intensive and costly. Insertion-null alleles (INNULs aims to evaluate the performance of the InnoTyper® 21 kit on a large set of highly degraded skeletal remains and compare its efficiency to the autosomal STR kit GlobalFiler™, considering the discriminatory power of both kits.

Material studied, methods, techniques

Material

This study analyzed 70 skeletal remains (Figure 1) recovered from mass graves in Andalusia. These samples were chosen for comparison due to partial or negative profiles obtained with GlobalFiler™. The remains had been buried at a depth of 3–4 meters for 70–80 years in an area known for high summer temperatures, low rainfall, and slightly acidic soil.

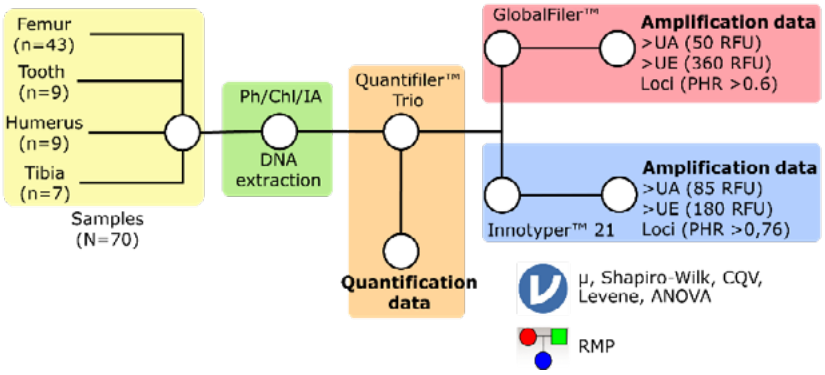


Figure 1. Material and methods

Methods

The samples were sanded with a Dremel® tool, cut into fragments, and exposed to UV light for 10 minutes per side. Tooth and bone fragments were pulverized using a TissueLyser II (QIAGEN, Hilden, Germany), and the resulting powder was transferred to a Falcon tube.

DNA was extracted using an organic extraction protocol, chosen for its high yield [3]as it requires efficient high-throughput methods. While little research has compared different techniques, silica in suspension has been identified in the literature as the best method for recovering small fragments, which are often present in these types of samples. In this study, we tested five DNA extraction protocols on 25 different degraded skeletal remains. Including the humerus, ulna, tibia, femur, and petrous bone. The five protocols were organic extraction by phenol/chloroform/isoamyl alcohol, silica in suspension, High Pure Nucleic Acid Large Volume silica columns (Roche. One gram of powder was mixed with lysis buffer and incubated overnight at 56°C. The lysate was mixed with phenol/chloroform/isoamyl alcohol, and the supernatant was concentrated and purified using the MinElute® PCR Purification Kit.

The purified extracts were quantified using the Quantifiler™ Trio kit and amplified with the GlobalFiler™ and InnoTyper® 21 kits. The same amount of DNA extract was used for each reaction.

The amplified samples were analyzed with a 3500 Genetic Analyzer, and the data was processed using GeneMapper™ IDX v1.6. Three parameters were analyzed: number of detected alleles, average RFU, and number of reportable loci. Thresholds were set according to SWGDAM guidelines.

Statistical analysis was performed using jamovi 2.2.5. Random match probability (RMP) was calculated using Familias version 3, with allele frequencies retrieved from 2023 GHEP-ISFG STRs and a previous study on InnoTyper® 21 in the Andalusian population [4]

Results

InnoTyper® 21 surpasses GlobalFiler™ in both the number of detected alleles and RFU, achieving nearly double in each case. Regarding reportable loci, InnoTyper® 21 tripled the count. Among the 22 samples that failed to produce reportable markers, 15 yielded positive results with InnoTyper® 21, though mostly as partial profiles. In instances where GlobalFiler™ detected 5 or fewer markers, InnoTyper® 21 often managed to obtain at least half of the profile. Statistical analysis using Shapiro–Wilk and Levene tests, followed by One-Way ANOVA, showed significant differences between the two kits in terms of the number of detected alleles (p-value < 0.001), average RFU (p-value = 0.003), and the number of reportable loci (p-value < 0.001). No significant differences were found in peak height ratio (p-value = 0.808). Comparing the results shown in **Figure 2**, it is clear that InnoTyper® 21 produces more markers. However, its discriminatory power is comparable to that of GlobalFiler™, with no significant differences in likelihood ratios (p-value = 0.321) according to One-Way ANOVA.

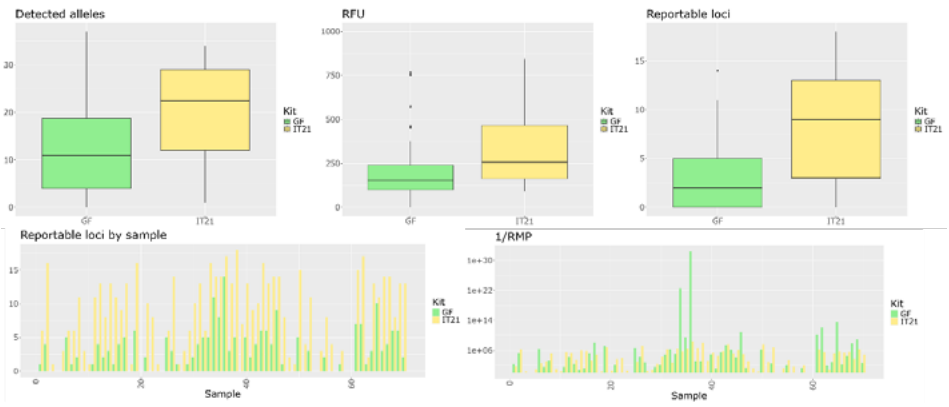


Figure 2. Key metrics for each commercial kit, including the number of detected alleles, RFU (Relative Fluorescence Units), and the number of reportable loci. The distribution of reportable loci per sample and the logarithm of the random match probability for each sample are also shown.

Discussion

Previous studies have shown that InnoTyper® 21 detects more alleles than autosomal STR kits like NGM™ [5] and GlobalFiler™ [6–8], resulting in more informative profiles, even when GlobalFiler™ yields zero markers. Our findings are consistent with these results, as InnoTyper® 21 detected more alleles and reportable markers. However, the profiles from skeletal remains with prior negative GlobalFiler™ results were not as informative as those in the study by Martins *et al.*, likely due to their samples being rootless hairs.

InnoGenomics tested InnoTyper® 21 with sonicated samples and found that higher random match probability values were achieved when DNA fragments were less than 150 bp, compared to GlobalFiler™ [9]. Our research supports this, showing that InnoTyper® 21 outperforms GlobalFiler™ in random match probability when fewer than 5 markers are detected with the autosomal STR kit. However, GlobalFiler™ has a much higher power of discrimination when 10 or more markers are detected.

This is why many authors suggest InnoTyper® 21 as a valuable complement to autosomal STRs [5,6,10]. Some even propose that InnoTyper® 21 offers higher discriminatory power than mitochondrial DNA [7], especially when few markers or a negative profile are obtained with autosomal STRs.

A final point to consider is whether the likelihood ratios from GlobalFiler™ and InnoTyper® 21 can be combined. This topic is widely debated, with some advocating against combining different DNA evidence [11] and others supporting it by using the product rule [12]. The key factor is demonstrating the independence between autosomal STRs and INNULs, which remains unresolved in the literature. However, a roughly calculated Kosambi recombination fraction [13] suggests extremely low values for markers located in the same loci: AC4027 and D7S820, TARBP and D1S1656, and NBC106 and FGA.

Conclusion

This research aimed to assess the effectiveness of InnoTyper® 21 for analyzing DNA from degraded skeletal remains and to compare the discriminatory power of partial INNUL profiles with that of autosomal STRs. Seventy degraded skeletal remains samples, which had yielded negative or partial profiles with GlobalFiler™, were analyzed using InnoTyper® 21. InnoTyper® 21 consistently produced more

alleles, higher RFU values, and more reportable loci than GlobalFiler™ in every sample. Despite these differences, the random match probability values from both commercial kits were similar. In conclusion, InnoTyper® 21 proves to be a valuable complementary tool for obtaining results from challenging samples that yield partial or negative profiles.

Acknowledgments

C.H. acknowledges the Spanish Ministry of Universities for funding the development of his PhD (FPU 20/01967).

And to Ms. Irene Izquierdo La Serna, for her invaluable support.

Conflict of interest statement

The authors declare no conflict of interest.

References

[1] C. Haarkötter, M. Saiz, X. Gálvez, D.C. Vinueza-Espinosa, M.I. Medina-Lozano, J.C. Álvarez, J.A. Lorente, Evaluation of the usefulness of insertion-null markers in critical skeletal remains, Int J Legal Med 138 (2024) 1287–1293. https://doi.org/10.1007/s00414-024-03205-3.

[2] J. Dabney, M. Meyer, S. Pääbo, Ancient DNA Damage, Cold Spring Harbor Perspectives in Biology 5 (2013) a012567.

[3] C. Haarkötter, X. Gálvez, D.C. Vinueza-Espinosa, M.I. Medina-Lozano, M. Saiz, J.A. Lorente, J.C. Álvarez, A comparison of five DNA extraction methods from degraded human skeletal remains, Forensic Science International 348 (2023) 111730. https://doi.org/10.1016/j.forsciint.2023.111730.

[4] M. Saiz, A. Gutierrez-Carvajal, X. Galvez, J.A. Lorente, J.C. Alvarez, Population genetic data of 20 insertion-null allele markers in the Andalusian population, International Journal of Legal Medicine 133 (2019) 1397–1400. https://doi.org/10.1007/s00414-019-02044-x.

[5] M. Van Den Berge, D. Wiskerke, R.R.R. Gerretsen, J. Tabak, T. Sijen, DNA and RNA profiling of excavated human remains with varying postmortem intervals,

Int J Legal Med 130 (2016) 1471–1480. https://doi.org/10.1007/s00414-016-1438-9.

[6] H. Brown, R. Thompson, G. Murphy, D. Peters, B. La Rue, J. King, A.H. Montgomery, M. Carroll, J. Baus, S. Sinha, F.R. Wendt, B. Song, R. Chakraborty, B. Budowle, S.K. Sinha, Development and validation of a novel multiplexed DNA analysis system, InnoTyper® 21, Forensic Science International: Genetics 29 (2017) 80–99. https://doi.org/10.1016/j.fsigen.2017.03.017.

[7] C. Martins, P.M. Ferreira, R. Carvalho, S.C. Costa, C. Farinha, L. Azevedo, António Amorim, M. Oliveira, Evaluation of InnoQuant® HY and InnoTyper® 21 kits in the DNA analysis of rootless hair samples, Forensic Science International: Genetics 39 (2019) 61–65. https://doi.org/10.1016/j.fsigen.2018.12.005.

[8] C. Martins, P.M. Ferreira, R. Carvalho, S.C. Costa, C. Farinha, L. Azevedo, A. Amorim, M. Oliveira, Internal validation of two new retrotransposons-based kits (InnoQuant® HY and InnoTyper® 21) at a forensic lab, Forensic Science International 283 (2018) 1–8. https://doi.org/10.1016/j.forsciint.2017.11.037.

[9] H. Brown, G.P. Murphy, A.H. Montgomery, J. Tabak, S.K. Sinha, Developmental validation of InnoTyper® 21, a nuclear DNA typing system based on retrotransposable element polymorphisms for degraded forensic samples, International Symposium on Human Identification (ISHI) (2016). https://innogenomics.com/wp-content/uploads/2021/01/Innogenomics_ISHI_2016_IT21_Dev_Valid_Poster_Final.cleaned.pdf.

[10] P.G. Ristow, N. Barnes, G.P. Murphy, H. Brown, K.W. Cloete, M.E. D'Amato, Evaluation of the InnoTyper® 21 genotyping kit in multi-ethnic populations, Forensic Science International: Genetics 30 (2017) 43–50. https://doi.org/10.1016/j.fsigen.2017.06.002.

[11] A. Amorim, A cautionary note on the evaluation of genetic evidence from uniparentally transmitted markers, Forensic Science International: Genetics 2 (2008) 376–378. https://doi.org/10.1016/j.fsigen.2008.04.001.

[12] B. Walsh, A.J. Redd, M.F. Hammer, Joint match probabilities for Y chromosomal and autosomal markers, Forensic Science International 174 (2008) 234–238. https://doi.org/10.1016/j.forsciint.2007.03.014.

[13] Fondevila, M., Santos, C., Phillips, C., Carracedo, A., Butler, J.M., Lareu, M.V., Vallone, P.M., An assessment of linkage between forensic markers: core STRs, mini-STRs and InDels, 22nd International Symposium on Human Identification (2011). https://strbase-archive.nist.gov/pub_pres/ISHI2011_Fondevila.pdf.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 93-99
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Enhanced Ancient DNA Extraction for Capillary Electrophoresis Analysis of Human Remains

Guihong Liu[1], Shengqiu Qu[1], Qiushuo Wu, Yazi Zheng, Mengyu Tan, Jiaming Xue, Meili Lv, Xiameng Chen, Lin Zhang*, Weibo Liang*

Department of Forensic Genetics, West China School of Basic Medical Sciences and Forensic Medicine, Sichuan University, Chengdu, Sichuan, China

[1] *These authors contributed equally to this work*

**Corresponding authors: Fax number: +86 28 85401825*

E-mail address: zhanglin@scu.edu.cn (L. Zhang), liangweibo@scu.edu.cn (W. Liang)

Abstract

The successful extraction of endogenous DNA from sclerous tissues is pivotal for the identification of human remains, but the retrieval of DNA from bone and tooth presents greater challenges compared to that from fresh tissues or bodily fluids. Existing methods or commercial kits for bone and tooth DNA extraction in forensic often struggle with aged, poorly preserved, chemically treated, or heat-exposed samples. Ancient DNA offer advanced extraction methods capable of isolating DNA from bone remains or sediments tens of thousands of years old, primarily utilizing silica adsorption. The samples in the two fields have different characteristics and the downstream analysis methods are not consistent, so the direct application of ancient DNA extraction methods to forensic may not produce the best results. This study used case and simulated samples to improve the ancient DNA extraction protocol to form a process that is more suitable for human remains analysis based on capillary electrophoresis platform. All in all, the research proves that the improved extraction methods of ancient DNA is more suitable for extracting highly degraded human remains in forensic, which helps to augment the success rate of human remains analyses.

Keywords

DNA extraction methods, human remains, ancient DNA.

Introduction

Human hard tissues, primarily bones and teeth, offer a certain level of protection for DNA, making them crucial sources for DNA extraction in individual identification and kinship analysis[1, 2]. These tissues play a pivotal role in scenarios involving the identification of human remains, such as homicide investigations, mass disasters, and wartime casualties[3]. However, the extraction of DNA from bones (or teeth) is notably more challenging compared to DNA retrieval from fresh tissues or body fluids[4, 5].

Currently, there is no universally recognized optimal method for bone DNA extraction[6, 7]. Ancient DNA (aDNA) extraction techniques have significantly advanced the field by enabling the retrieval of DNA from skeletal remains or sediments that date back tens of millennia to several hundred thousand years. Among these techniques, DNA extraction methods based on silica adsorption have gained widespread use[8]. The most commonly employed aDNA extraction protocol was introduced by Dabney et al.[9], which, when combined with single-stranded library preparation, allows the recovery of DNA fragments as short as 35 base pairs (bp). Due to its superior ability to retain short fragments, this method is particularly well-suited for samples exhibiting extensive DNA degradation. While initially developed for aDNA, this extraction method has recently been adopted in forensic science and has demonstrated advantages in processing highly degraded forensic samples[10, 11].

However, the downstream analytical workflow used by aDNA extraction techniques typically involves direct sequencing[12], which is different from the short tandem repeat (STR) analysis that utilizes multiplex amplification and capillary electrophoresis commonly used in forensic investigations[13]. Moreover, the chronological age of the samples in aDNA studies tends to be significantly older[14]. Thus, the direct application of aDNA extraction protocols to forensic samples may not always be optimal. This study aims to comprehensively compare the performance of aDNA extraction methods with forensic protocols for human remains. Additionally, we adapted the aDNA extraction protocol for forensic applications, resulting in an optimized extraction method. The potential of this optimized protocol in forensic casework was then initially evaluated using case samples.

Materials and methods

The femoral samples used in this study were obtained from the West China Forensic Identification Center of Sichuan University and were properly preserved under laboratory conditions prior to processing. The samples underwent sequential cleaning with hypochlorous acid, deionized water, and ethanol, followed by overnight drying in a laminar flow cabinet. Bone powder was collected from the mid-diaphysis using an electric drill. DNA extraction was then performed on the bone powder using both a forensic bone extraction method based on a centrifugal column and an aDNA extraction protocol. Each extraction experiment was performed on the same sample three times, and the final result was the average of the three repeated extraction experiments.

For forensic bone extraction method, bone powder was lysed overnight with a bone lysis buffer (Promega, USA), after which the lysate was combined with PB buffer (Qiagen, Germany) and processed through a centrifugal column (Qiagen, Germany), followed by elution with TE buffer. Two aDNA extraction protocols were utilized: the centrifugal column method and the magnetic bead method. Both protocols shared the same lysis procedure, employing a lower lysis temperature and a simplified lysis buffer (primarily containing proteinase K and ethylenediaminetetraacetic acid) for overnight incubation. In the aDNA centrifugal column method, a large volume of a self-prepared guanidine hydrochloride buffer was combined with the lysis buffer and passed through a centrifugal column (Qiagen, Germany), followed by elution with TET buffer. In the aDNA magnetic bead method, a self-prepared guanidine hydrochloride buffer, magnetic bead suspension (VWR, USA), and lysis buffer were used to bind the DNA to magnetic beads, followed by elution with TET buffer.

Subsequent improvements and optimizations were made to the aDNA extraction protocols, specifically targeting the lysis and purification steps. These optimizations included adjustments to lysis temperature, duration, buffer composition, purification methods, and the ratio of binding solutions. A final optimized extraction protocol was developed, and the previously described samples were re-extracted using this improved method.

Following extraction, the AGCU Expressmarker 22 fluorescence detection kit was employed to perform multiplex amplification of short tandem repeat (STR) loci for all DNA samples. The PCR products were subjected to capillary electrophoresis on an ABI 3500 genetic analyzer, and the resulting data were analyzed using GeneMapper software. STR profiles were compared based on three criteria: average peak height, number of amplified loci, and number of alleles detected.

Results

Figure 1 presents the results of three replicate extractions performed on the same femoral sample using three different extraction methods. Among the methods, the aDNA magnetic bead method yielded the highest average peak height, number of amplified loci, and number of alleles. Notably, the difference in the number of amplified loci between the aDNA magnetic bead method and the forensic bone extraction method was statistically significant ($P < 0.05$). Additionally, the aDNA centrifugal column method produced slightly higher values for average peak height, number of amplified loci, and number of alleles compared to the forensic bone extraction method.

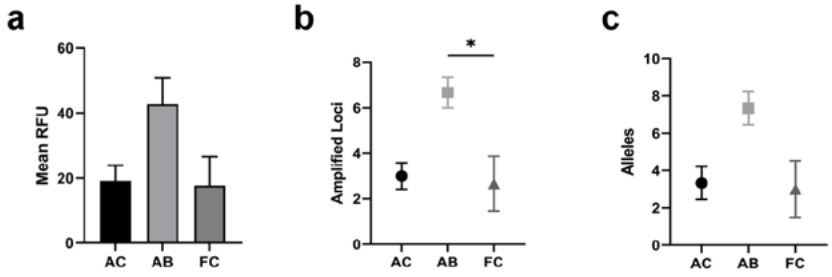


Figure 1. The STR profiles of the same sample were obtained through three repeated extractions using three different methods. AC represents the aDNA centrifugal column method, AB denotes the aDNA magnetic bead method, and FC corresponds to the forensic bone extraction method.

Figure 2 illustrates the comparison between the optimized aDNA extraction method and the forensic bone extraction method. The optimized protocol demonstrated superior performance in terms of average peak height, number of amplified loci, and number of alleles. The differences in the number of amplified loci and alleles between the two methods were statistically significant ($P < 0.05$).

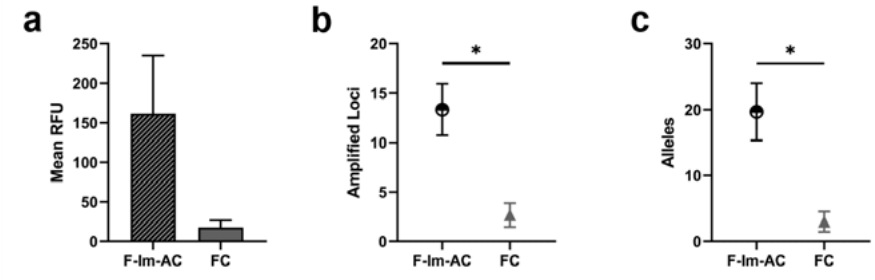


Figure 2. The STR profiles of the same sample were obtained through three repeated extractions using two different methods. F-Im-AC represents the optimized extraction method, and FC corresponds to the forensic bone extraction method.

Discussion

The initial comparison indicates that the aDNA *extraction method has certain advantages. To further enhance its extraction efficiency, we optimized this protocol, with details of the optimization process to be published in subsequent articles. The optimized method demonstrated improved performance, significantly increasing the average peak height, number of amplified loci, and number of alleles. Currently, the samples used in this study are relatively homogeneous. Expanding the research to include samples from various time periods, storage conditions, and types would broaden the applicability of this method.*

Conclusion

aDNA extraction methods demonstrate a slight advantage over forensic bone extraction methods, with these advantages becoming more pronounced following optimization. These improvements contribute to a higher success rate in the analysis of human remains.

Acknowledgments

The authors would like to thank all the participants who volunteered for this study. This work was supported by the National Natural Science Foundation of China (No. 82271929), the Sichuan University Archaeological Science Center Open Project (No. 23SASB04), and Open Project of Sichuan University Key Laboratory of Criminal Investigation (No. 2024ZD05).

Conflict of interest statement

The authors have no competing interests to declare.

References

[1] Raffone C, Baeta M, Lambacher N, et al. Intrinsic and extrinsic factors that may influence DNA preservation in skeletal remains: A review[J]. Forensic Sci Int, 2021,325: 110859.

[2] Alaeddini R, Walsh S J, Abbas A. Forensic implications of genetic analyses from degraded DNA--a review[J]. Forensic Sci Int Genet, 2010,4(3): 148-157.

[3] Raimann P E, Picanco J B, Silva D S, et al. Procedures to recover DNA from pre-molar and molar teeth of decomposed cadavers with different post-mortem intervals[J]. Arch Oral Biol, 2012,57(11): 1459-1466.

[4] Latham K E, Miller J J. Dna recovery and analysis from skeletal material in modern forensic contexts[J]. Forensic Sci Res, 2019,4(1): 51-59.

[5] Hasap L, Chotigeat W, Pradutkanchana J, et al. A novel, 4-h DNA extraction method for STR typing of casework bone samples[J]. Int J Legal Med, 2020,134(2): 461-471.

[6] Heathfield L J, Haikney T E, Mole C G, et al. Forensic human identification: Investigation into tooth morphotype and DNA extraction methods from teeth[J]. Sci Justice, 2021,61(4): 339-344.

[7] Finaughty C, Heathfield L J, Kemp V, et al. Forensic DNA extraction methods for human hard tissue: A systematic literature review and meta-analysis of technologies and sample type[J]. Forensic Sci Int Genet, 2023,63: 102818.

[8] Liu Y, Mao X, Krause J, et al. Insights into human history from the first decade of ancient human genomics[J]. Science, 2021,373(6562): 1479-1484.

[9] Dabney J, Knapp M, Glocke I, et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments[J]. Proc Natl Acad Sci U S A, 2013,110(39): 15758-15763.

[10] Xavier C, Eduardoff M, Bertoglio B, et al. Evaluation of DNA Extraction Methods Developed for Forensic and Ancient DNA Applications Using Bone Samples of Different Age[J]. Genes (Basel), 2021,12(2).

[11] Emery M V, Bolhofner K, Winingear S, et al. Reconstructing full and partial STR profiles from severely burned human remains using comparative ancient and forensic DNA extraction techniques[J]. Forensic Sci Int Genet, 2020,46: 102272.

[12] Orlando L, Ginolhac A, Zhang G, et al. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse[J]. Nature, 2013,499(7456): 74-78.

[13] Rucinski C, Malaver A L, Yunis E J, et al. Comparison of two methods for isolating DNA from human skeletal remains for STR analysis[J]. J Forensic Sci, 2012,57(3): 706-712.

[14] Hofreiter M, Sneberger J, Pospisek M, et al. Progress in forensic bone DNA analysis: Lessons learned from ancient DNA[J]. Forensic Sci Int Genet, 2021,54: 102538.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 101-106
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Longitudinal assessment of DNA recovery from teeth recovered from a semi-natural marine environment

L. Malan[a] V.E. Gibbon[b], C. Finaughty[a], L.J. Heathfield[a]

[a]*Division of Forensic Medicine and Toxicology, Department of Pathology, Faculty of Health Science, University of Cape Town, South Africa*

[b] *Division of Clinical Anatomy and Biological Anthropology, Department of Human Biology, University of Cape Town, South Africa.*

Abstract

Teeth are a well-known source of DNA for forensic human identification. However, there is still a lack of experimental research that investigates DNA recovery over time in realistic marine conditions. In this study, sixteen pairs of donated human wisdom teeth were exposed to a semi-natural marine environment (tanks) that simulated seawater conditions of Cape Town, South Africa. Seawater tank temperature were maintained between 10.1-14.5°C, with sampling occurring over 30 days. DNA quantification and profiling, in addition to massively parallel sequencing, was done to assess DNA recovery over time. The statistical informativeness of the forensic DNA profiles was used to assess DNA recovery. DNA concentration declined over time for both control and experimental teeth, and forensic DNA profiles were only partially informative by 30 days. The study provided valuable insight into DNA recovery from marine environments, which can improve forensic DNA identification of human remains retrieved from marine environments.

Key words

DNA profiling, forensics genetics, hard tissue, teeth, ocean, capillary electrophoresis.

Introduction

Teeth are valuable sources of DNA for forensic human identification purposes. Although teeth are often used for highly degraded, or heat altered human remains, there is limited empirical, realistic experimental research investigating DNA recovered from teeth submerged in the ocean. Existing studies have focused on case reports, with few comparing submersion duration [1–4]. Therefore, DNA recovery

from donated human wisdom teeth exposed to experimental seawater tanks, simulating semi-natural marine environment conditions of the western coast of Cape Town, South Africa.

Materials and methods:

Sixteen pairs of teeth were collected from consenting patients undergoing wisdom tooth removal surgery. The pairs were split into one control and one experimental third molar. Teeth with visible pulp, dental diseases, or fragmented conditions were excluded. Holes were drilled in centrifuge tubes, each containing one experimental tooth, and submerged in a seawater tank at a depth of around 30 cm. The tanks were filled with natural seawater, and temperatures were maintained between 10.1°C-14.5°C. Teeth were sampled everyday for the first 10 days, with further samplings on days 12, 14, 16, 18, 20, and 30 thereafter. Control teeth were stored at room temperature in a dry laboratory for the same period.

All teeth were powdered using a TissueLyser and 0.05g powder underwent demineralisation in combination with a cell lysis solution at 56°C for 20 hours. The QIAamp® DNA Investigator Kit was used for DNA purification. DNA quantification was carried out using real-time PCR with the Applied Biosystems Quantifier Trio Kit, and analysed using HID Real-Time PCR Analysis Software. The degradation index (DI) was calculated and interpreted as previously described [4].

The GlobalFiler™ PCR Amplification Kit was used for forensic DNA profiling. Thermal cycling was performed using Bio-Rad's T100 Thermal Cycler, and capillary electrophoresis (CE) was conducted on the 3500xl Genetic Analyzer for Human Identification. Data from CE were analysed using GeneMapper ID-X software. Profiling success was categorised based on statistical informativeness, overcoming limitations of standard profiling terms.

The samples from day 20 underwent massively parallel sequencing (MPS) with the ForenSeq™ DNA Signature Prep Kit (Verogen™), following the manufacturer protocol, except. DNA concentrations were below 0.2 ng/µL and thus required 5 µL input DNA for library preparation. Libraries were purified, normalised, and pooled before sequencing on the MiSeq FGx System.

Results

For both experimental teeth (*p<0.001, R=-0.804*) and non-submerged controls (*p<0.001, R=-0.716*), strong negative exponential correlations were observed between average DNA concentration and time (Figure 1a). A consistent decrease in DNA concentration was noted with increasing submersion duration, both in the experimental and matched non-submerged controls, with concentrations ranging from 0.37-0.001 ng/mg and 0.52-0.01 ng/mg, respectively (Figure 1a).

No correlation was found between DI and time for both experimental teeth (*p=0.56*, R=-0.156) and matched non-submerged controls (*p=0.53*, R=0.171) (Figure 1b), with both having mean DIs less than 1.5, indicating no considerable DNA degradation. No significant differences in concentration and DI were observed between matched non-submerged controls and experimental teeth.

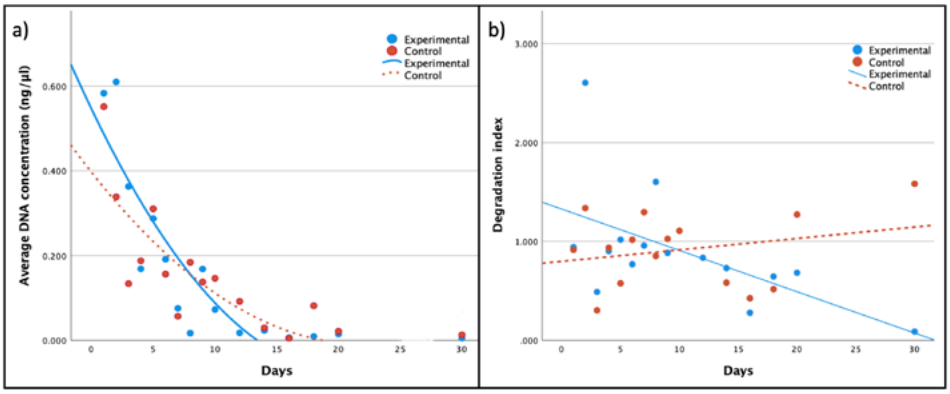


Figure 1a) Line graphs of the a) average DNA concentration (ng/mg) and **b)** degradation index over days for non-submerged control teeth and experimental samples.

Extended submersion led to decreased forensic DNA profile quality. Statistically informative profiles were obtained for experimental teeth until day 16, followed by partially informative on days 20 and 30 (Figure 2). Informative DNA profiles were obtained for the matched non-submerged controls on days 20 and 30. Although a decrease in DNA profiling success was observed after day 16, there was no statistical correlation between DNA profiling success and submersion period (R=-0.130, *p=0.239*). No significant differences were observed between the profiling success rate of the experimental teeth and matched non-submerged controls (Z=-0.593,

p=0.553). When MPS was conducted on the day 20 sample, 24 STRs and five SNPs were sequenced, compared to the 17 alleles successfully sequenced during CE, with concordance on all overlapping markers.

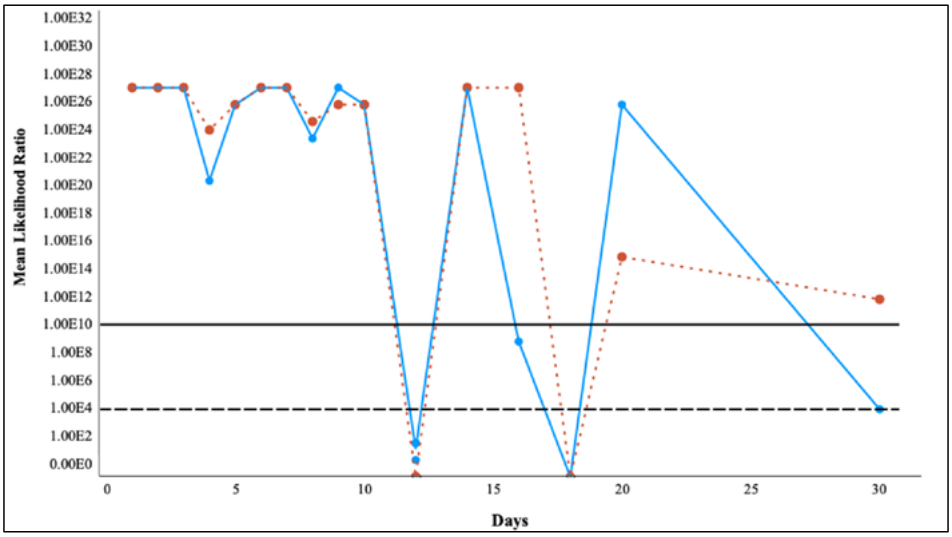


Figure 2. Scatter plot of the likelihood ratios of teeth samples over time. Above A is statistically informative, between A and B is partially informative, and below B is Statistically uninformative. Controls = Matched non-submerged controls.

Discussion:

On average, submerged teeth had lower statistical identification power than the matched non-submerged controls. As demonstrated by other studies, forensic DNA profiling success is reduced with increased submersion time in seawater [3]. DNA profiling success was relatively low compared to results from other studies on teeth from terrestrial environments or storage at low temperatures [5]. Failed DNA profiles were observed for both the matched non-submerged controls and the experimental samples for days 12 and 18. The reason for these failed profiles might not be attributed to seawater solely, as full DNA profiles were seen for other samples for longer exposure times, thus suggesting that poor DNA recovery may be due to intrinsic factors [5,6].

A noteworthy finding was that DNA degradation, measured as DI, stayed relatively low and constant throughout the deployment. Few studies consistently

reported the DI and PCR inhibition in water-logged samples, rather focussing on DNA concentration only [3]. However, in this study, as seen with others, DNA concentration does not accurately predict DNA profiling success [7]. In contrast, several studies have concluded that DI was indicative of the DNA profile quality and may inform sample processing procedures to ensure maximum allele recovery [8].

Statistically informative DNA profiles were obtained up until day 20. Furthermore, depending on the DNA platform used, more informative profiles could be obtained. For example, for MPS, more alleles were amplified compared to CE and they yielded a higher likelihood ratio for the day 20 sample. A previous case study on human bone samples recovered from Table Bay after two months successfully produced full A-STR, X-STR, and Y-STR profiles [1]. Teeth provide a protective barrier from environment exposure, and DNA integrity may be due to the chemical and anatomical composition of teeth preventing exogenous contamination and degradation; the salt in ocean marine water might also may preserve the DNA [2,9]; or the low water temperatures may preserve the DNA [10].

Conclusions

This study represents the first of its kind where DNA recovery from human teeth are exposed longitudinally to realistic seawater conditions. There was variability observed with forensic DNA profiling success, but statistically informative profiles were still obtainable for later submersion times. In particular, DNA concentration is not a good predictor for DNA profiling success.

Conflict of interest:

No authors have any conflicts of interest to declare.

Acknowledgements

We hereby acknowledge the Division of Forensic Medicine and Toxicology at UCT, the UCT Research Diving Unit, Ms Andrea Plos, Dr Ranchod and Dr Greef for their assistance with sample collection, and all the donors for this study. This project was financially supported by the South African National Research Foundation Thuthuka

Grant to LJH (Grant No. 129499); VG acknowledges (Grant No. 115357) and their financial assistance towards this research is hereby acknowledged. The conclusions arrived at are those of the authors and are not necessarily attributed to the NRF.

References:

[1] Finaughty C, Reid KM, Alli IH, Heathfield LJ. A first for forensic genetics in Africa: successful identification of skeletal remains from the marine environment using massively parallel sequencing. Forensic Sci Int Genet 2020;49:102370.

[2] de Oliveira Musse J, da Costa Nardis A, Anzai EK, Hirata MH, Barreto Cicarelli RM, Nogueira de Oliveira R. Freshwater and salt-water influence in human identification by analysis of DNA: an epidemiologic and laboratory study. Braz J Oral Sci. 2009;8.

[3] Hughes-Stamm S, Warnke F, van Daal A. An alternate method for extracting DNA from environmentally challenged teeth for improved DNA analysis. Leg Med 2016;18:31–6.

[4] Finaughty C, Gibbon VE, Speed B, Heathfield LJ. Lost at sea: A pilot study investigating DNA recovery from teeth in a South African natural marine environment. Forensic Sci Int Genet Suppl Ser 2019;7:580–1.

[5] Simoes Dutra Correa H, Cortellini V, Brescia G, Verzeletti A. Human identification through DNA analysis of restored postmortem teeth. Forensic Sci Int Genet 2020;47:102302.

[6] Raffone C, Baeta M, Lambacher N, Granizo-Rodríguez E, Etxeberria F, de Pancorbo MM. Intrinsic and extrinsic factors that may influence DNA preservation in skeletal remains: A review. Forensic Sci Int 2021;325:110859.

[7] Budowle B, Eisenberg AJ, Van Daal A. Validity of Low Copy Number Typing and Applications to Forensic Science. Croat Med J 2009;50:207.

[8] Vernarecci S, Ottaviani E, Agostino A, Mei E, Calandro L, Montagna P. Quantifiler ® Trio Kit and forensic samples management: A matter of degradation. Forensic Sci Int Genet 2015;16:77–85.

[9] Singh A, Singh N. Effect of salt concentration on the stability of heterogeneous DNA. Physica A: Statistical Mechanics and Its Applications 2015;419:328–34.

[10] Lozano-Peral D, Rubio L, Santos I, Gaitán MJ, Viguera E, Martín-de-las-Heras S. DNA degradation in human teeth exposed to thermal stress. Sci Rep 2021;11:12118.

30th INTERNATIONAL SOCIETY FOR FORENSIC GENETICS CONGRESS
Universidade de Santiago de Compostela, 2025, pp. 107-114
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Molecular analysis of ancient human remains belonging to multiple unknown burials discovered in an Italian middle age archaeological site

Denise Gianfreda[1*], Beatrice Corradini[1], Francesca Ferrari[2], Ilaria Borciani[1], Laura Donato[3], Anna Laura Santunione[1], Gianmarco Ferri[1], Rossana Cecchi[1]

[1] *Department of Biomedical, Metabolic and Neural Sciences, Institute of Legal Medicine, University of Modena and Reggio Emilia, Italy.*
[2] *University Hospital of Modena, Italy.*
[3] *Institute of Legal Medicine, University of Tor Vergata, Italy.*

** Corresponding author at: denise.gianfreda@unimore.it.*

Abstract

In forensics, the study of archaeological sites is of considerable importance to understand the dynamics of burials, rule out kinship scenarios and unravel historical questions. In this study, molecular analyses were conducted on skeletal remains of belonging to a burial site found in the area beneath the cloister outside the former convent of S. Francesco in Mirandola (Modena, Italy). During excavations, the portion of a late-medieval burial ground was discovered, with infant and adult graves dating back to a period between the 13th and 16th centuries. Nearly thirty burials were identified, placed in pits dug into the ground, organized on several overlapping levels and separated only by a couple of centimetres of soil, oriented according to the Christian tradition along an East-West axis. The historical explanation for this discovery is currently poorly known and no archive records are available. This analysis aims to integrate the archaeological and anthropological studies already underway on the site, reveal any biological relationships among the individuals, confirm the exact number of subjects and explaining their peculiar arrangement and location, finally reconstruct appearance and population of origin. These findings could be helpful to deep the knowledge about the burial site and the city's history during the medieval period. Moreover, the performance of routinely adopted laboratory workflow applied to these highly compromised remains was assessed. A preliminary analysis was conducted on one molar tooth and, after the promising results, the analyses were extended to the entire group of remains. After decalcification, genomic DNA was isolated using a powder-free extraction method from bones and dental samples. Autosomal and Y-STR markers were genotyped, additionally HIrisPlex system was used to predict eye and hair colour. To date, DNA data were obtained from ten of the eleven individuals analysed, further analyses will be carried out to provide genetic results for the entire burial site.

Keywords

Human identification, ancient DNA, kinship analysis, DNA typing, forensic DNA phenotyping, medieval burial site.

1. Introduction

A molecular study was conducted on human skeletal remains belonging to a burial site found in the area beneath the cloister outside the former convent of S. Francesco in Mirandola (Modena, Italy). During the excavation works set up to repair earthquake's damages, several burials were found containing infant, sub-adult and adult skeletal remains dating back around the 13th and 14th centuries AD. It's supposed that the necropolis was set up during the first construction of the church occurred at that time, however it's known that the area was subjected to secondary alteration until the 16th century [1]. Contextual informations are scarce and incomplete, no archive records are available about the identity, origin and period in which the individuals lived, leading to a lack of historical explanation for this discovery.

Preliminary archaeological and anthropological studies reveal the presence of thirty-two burials of fourteen adults, nine subadults, eight infants, plus scattered bones of which some attributable to secondary depositions. The discovered remains were overall highly compromised, only ten skeletons were almost complete and other ten partially preserved, for twelve graves only few bones were founded. Morphological sex diagnosis identifies four females and seven males, for the remaining twenty individuals, sex was undetermined. The graves were discovered at approximately 1.5 m depth from the walking surface placed in pits dug into the ground, organized on several overlapping levels and separated only by a couple of centimetres of soil. Arrangements of bodies follow the Christian tradition: tombs were single graves, the inhumed were interred directly into the ground and arranged in an orderly way, vertically, along an East-West axis in the direction of the rising sun, without a funeral equipment and allegedly wrapped in a shroud. Some exceptions occurred: two burials were double graves; the double burial T8 was North-South oriented with skeletons partially overlaid and skulls rotated towards each other; the burial T25 was the only one surrounded by a brick crate [1].

This genetic study aims to integrate the archaeological and anthropological studies already underway on the site, looking for kinships between the inhumates that could explain burial arrangements and location, confirm the exact number of subjects, associate scattered bone remains and reconstruct information about genetic ancestry and appearance.

2. Primary molecular analysis

Permission to access to the human remains and samples collection was granted by the Ministry of Culture, Archaeological Superintendence of Fine Arts and Landscape of the provinces of Bologna, Modena, Reggio Emilia and Ferrara. Most of the skeletal remains founded were in a poor state of preservation due to the time spent underground and the manner of deposition, suggesting difficulties in applying the routinely adopted laboratory workflow for the genetic analysis of highly compromised remains.

In order to assess the feasibility of the study, a preliminary analysis was conducted on a molar tooth belonging to individual T15. After the promising results, the analyses were extended to the other skeletal remains. Teeth and bones samples were collected based on the state of preservation of the bodies and the availability of bone material. For twelve adults and six subadults teeth were sampled, respectively permanent and deciduous, preferably molars. Additionally, femur and petrous bones were also collected from subadults and skull-less adults, based on the preservation of skeletal parts. To date, best-preserved teeth samples of eleven individuals were subjected to DNA extraction (Table 1). When available, DNA was extracted from at least two teeth samples per individual, otherwise DNA extraction was performed in replicates.

Teeth pre-treatment consisted of a cleaning procedure of the external surface carried out with distilled water and brushes to remove ground residues and exogenous material, a decontamination phase performed with bleach washes and finally a grinding operation was made by using a vise. A powder-free protocol was applied for demineralization process [2]. Fractured teeth were incubated with 0.5 M EDTA pH 8.0 at 56°C in water bath for about 10 days, with periodic changes. Complete demineralized pieces were cut with a scalpel into smaller chips used directly for DNA extraction. An Investigator Lyse&Spin Basket (Qiagen) paired with QIAamp DNA Investigator Kit (Qiagen) modified manufacturer's protocol for bones and teeth was applied, DNA extracted was eluted in a final volume of 25 µl of ATE buffer. DNA extracted quantity and quantity were checked by PowerQuant® System (Promega) on 7500 Real-Time PCR System (Applied Biosystems).

To perform sex diagnosis and preliminary kinship analysis, STRs typing of autosomal DNA were performed with PowerPlex® Fusion System (Promega), GlobalFiler™ IQC and VeriFiler™ Plus PCR Amplification Kits (Applied Biosystems). For males, also Y-STRs haplotypes were obtained using PowerPlex® Y23 System (Promega). PCR amplifications were performed in replicates on Veriti™

Thermal Cycler (Applied Biosystems), sequencing and data analysis were carried out on SeqStudio™ Genetic Analyzer for HID (Applied Biosystems). STR alleles calling was performed with GeneMapper™ ID-X 1.6 (Applied Biosystems) software applying a 50 RFU threshold.

The HIrisPlex system was used for prediction of eye and hair colour. The 24 SNPs targeted by the multiplex PCR were genotyped through the SNaPshot™ technique (Applied Biosystems) and GeneMapper™ 6 (Applied Biosystems), phenotyping predictions were undergone with the HIrisPlex system webtool [4].

The Familias software v3.4 was used for kinship analysis, Likelihood Ratio (LR) calculations were made using the Blind search tool implemented in the software [5] and the STRidER v2/R2 European dataset for autosomal STR allele frequencies [6]. Additionally, YHRD database's kinship analysis tool was utilized for Y-STRs haplotypes ancestor-offspring LR evaluations [7]. For relationship testing were taken into account the genetic profiles with a number of confirmed loci (amplified at least in two PCR replicates) above a threshold of 10 STRs [8], for STRs profiles obtained with different kits, only the shared loci were taken into account for the suitability establishment.

3. Preliminary results and discussion

DNA was successfully extracted from adults' permanent teeth samples, averages of autosomal DNA extracted are listed in Table 1. Quantity and quality of the genetic material, resulted from qPCR analysis, was in accordance with the state of preservation observed for each dental remains. The powder-free demineralization and extraction protocol applied thus allowed to obtain significant results, despite the highly unfavourable conditions of the samples analysed. Negative results were obtained from DNA extractions of T20 infant burials' deciduous teeth. However, as already pointed out in literature, deciduous teeth showed lower DNA yield, amount and preservation compared to adult permanent teeth [9]. For graves whose DNA extracted was poorly or highly degraded, petrous bone and femur analysis will be further developed.

STRs typing analysis results in full and partial profiles for samples whose qPCR DNA extract concentration measurement resulted of greater than 1 pg/µl. PowerQuant® Analysis Tool (Promega) output data were utilized to establish DNA input for PCR amplifications. The number of autosomal STRs loci recovered with regard to PCR DNA input was extremely variable, considering the three different

amplification kits used and the degradation degree of each sample. For ten of the eleven individuals tested, it was possible to obtain at least two replicates, that enable to confirm STRs loci for kinship analysis.

3.1. Sex determination comparison

Molecular sex determination was performed comparing Amelogenin and Y-STRs typing results. According to DNA results, five individuals were identified as females, four as males and two were undetermined (Table 1). For five individuals, morphological attribution was confirmed, resulting in correct morphological identification of two females and three males. However, for four individuals, morphological attribution was not confirmed, since genetic analysis identifies as females three individuals previously determined as males and one male previously assigned as a female. The concordance between morphological and genetic results are listed in Table 1. The infant of T20 burial, whose sex diagnosis could not be previously performed, is still undetermined to date and sex diagnosis confirmation for T24 was also not possible due the lack of genetic data for this two buried.

3.2. Phenotype prediction

Multiplex PCR amplifications of predictive 24 SNPs targeted by HIrisPlex system results in one full SNPs profile for the individual of the T3 burial, partial profiles were obtained for six individuals. In Table 1 is indicated the number of SNPs recovered for each buried analysed. In T24 and T4 profiles, the SNPs related to MC1R gene were the most affected by drop-out, as already advised by Walsh et al. [3]. The prediction analysis results are summarized in Table 1, the p-value corresponding to the highest predicted categories, the AUC loss and the predicted phenotype for eye and hair colour are also displayed. All eye colour prediction resulted above 0.7 p-value thresholds, except for T3 individual that results in a p-value of 0,668 for brown eye (AUC loss = 0) suggesting a lighter colour tending to hazel, considering a 0.5 p-value threshold [3,10]. Hair colour predictions (applying a p-value threshold of 0.7) result overall in a dark brown to black phenotype [3]. Preliminary appearance prediction of individuals analysed reveals brown eye and hair colour, common traits of the Italian population.

3.3. Kinship analysis

Autosomal and Y-STRs profiles were used for kinship investigation. Four individuals (T3, T4, T15 and T28) were suitable for relationship testing, having a number of confirmed autosomal STRs loci in the consensus profile above 10 (Table 1).

LR calculations results, arising from Familias Blind search tool, were all minor or close to 1, favouring hypothesis of individuals not being related for the degree of kinship tested (parent-child, siblings, half-siblings, cousins, 2nd-cousins). A single exception resulted for an obtained LR of 2.20, indicating a weakly supported cousins' relationship between T15 and T4 individuals. However, these results are preliminary and will be investigated further when complete data will be available.

Ancestor-offspring LR evaluations were performed to establish closely relationships through YHRD database kinship analysis tool. Four unique Y-STR haplotypes were observed, hence, YHRD kinship-index calculation resulted in a more likely patrilineal non-relationships for all kinship tested.

Considering the preliminary results, no evaluation of historical importance was undergone. However, the interesting data obtained allow for more promising in-depth further analyses to contextualise the human remains discovered in the medieval Italian historical scenario.

5. Acknowledgments

Authors express their gratitude to Carolina Ascari Raccagni (Archaeological Superintendence of Fine Arts and Landscape of the provinces of Bologna, Modena, Reggio Emilia and Ferrara, Bologna, Italy) and Simona Scaruffi (ArcheoModena Professional association Benassi Guandalini Scaruffi, Modena, Italy) for allowing this project. Authors thanks the graduate student Francesca Campagna for her effort.

6. Conflict of interest statement

Authors declare that they have no conflict of interest.

7. References

[1] Ascari Raccagni C., Preghiere semplici. Scavi archeologici e ricerche all'ex convento di San Francesco a Mirandola, Ante Quem, Bologna, 2024.

[2] Ferri G., Corradini B., Gianfreda D. et al. Two caseworks for one gene: successful species identification from compromised bone materials with the 12S rRNA.

Int J Legal Med 136, 1255–1260 (2022). https://doi.org/10.1007/s00414-022-02817-x.

[3] Walsh S., Chaitanya L., Clarisse L. et al. Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci Int Genet. Mar;9:150-61. (2014) https://doi.org/10.1016/j.fsigen.2013.12.006.

[4] https://hirisplex.erasmusmc.nl/.

[5] Kling D., Tillmar A.O., Egeland, T. Familias 3-Extensions and new functionality. Forensic Sci Int Genet. (2014) Nov;13:121-7. https://doi.org/10.1016/j.fsigen.2014.07.004.

[6] https://strider.online/frequencies.

[7] https://yhrd.org/.

[8] Raccomandazioni GE.F.I. nelle indagini di identificazione personale (2018) https://www.gefi-isfg.org/temp/20112018100445.pdf.

[9] Leskovar T. and Zupanič Pajnič I. Comparative analysis of DNA preservation in permanent and deciduous teeth of adults and non-adults: Implications for archaeological and forensic research. Forensic Sci Int. Dec;353:111882. (2023). https://doi.org/10.1016/j.forsciint.2023.111882.

[10] Walsh S., Lindenbergh A., Zuniga S.B. et al. Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence. Forensic Sci Int Genet. Nov;5(5):464-71. (2011) https://doi.org/10.1016/j.fsigen.2010.09.008.

8. Tables

Table 1: Summary of the analysed graves data [1] and genetic results. aMean of PowerQuant® system [AUTO] results. bRespectively, number of loci amplified with PowerPlex Fusion, GlobalFiler IQC, VeriFiler Plus Kits. cLoci considered in the consensus profile, amplified at least in two PCR replicates. dPowerPlex Y23 results. ePhenotypes for the highest probability value are abbreviated as Br for Brow, Bl for Blue in eye and Black in hair, D for Dark.

Graves ID	Morpho-logical age	Teeth samples analyzed	Mean DNA extracted [pg/µl][a]	Autosomal STRs		N° of Y-STRs typed[d]	Sex diagnosis		N° of SNPs amplified (HIrisPlex system)	Eye colour prediction			Hair colour prediction				
				N° of loci typed per kit[b]	N° of loci in consensus profile[c]		Morphology	Genetic		Highest colour p-value[e]	AUC loss	Predicted colour	Highest colour p-value[e]	AUC loss	Highest shade p-value[e]	AUC loss	Predicted colour
T2	Adult	Incisive Canine	8,87	16, -, -	9	-	Male	XX	-	-	-	-	-	-	-	-	-
T3	Adult	Molar Incisive	66,87	22, -, 23	22	23	Male	XY	24	0,668 *Br*	0	Brown to Hazel	0,676 *Br*	0	0,588 *D*	0	Dark brown
T4	Adult	Premolar	1,15	16, 18, 23	15	23	Male	XY	11	0,848 *Bl*	0	Blue	-	-	-	-	-
T6	Adult	Molar Premolar	2,75	21, 18, -	8	-	Female	XX	15	0,962 *Br*	0,015	Brown	-	-	-	-	-
T7	Adult	Molar Incisive	5,34	15, 18, -	8	-	Male	XX	19	0,985 *Br*	0,002	Brown	0,715 *Br*	0,006	0,840 *D*	0,004	Dark brown
T8a	Adult	Molar	15,0	19, 18, 20	9	12	Male	XY	-	-	-	-	-	-	-	-	-
T15	Suba-dult	Molar	5,75	15, 21, 22	18	-	Female	XX	22	0,986 *Br*	0	Brown	0,660 *Br*	0,004	0,677 *D*	0,004	Dark brown
T20	Infant	Molar Incisive canine	-	-, -, -	-	-	-	-	-	-	-	-	-	-	-	-	-
T24	Adult	Molar Incisive	-	18, 6, -,	0	-	Male	-	7	0,981 *Br*	0,013	Brown	-	-	-	-	-
T25	Adult	Molar Incisive	4.70	13, -, -	4	-	Male	XX	-	-	-	-	-	-	-	-	-
T28	Adult	Molar	-	16, 20, 23	14	18	Female	XY	12	0,969 *Br*	0,029	Brown	0,698 *Bl*	0,026	0,932 *D*	0,044	Dark black

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 115-125
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Optimisation and evaluation of forensic DNA recovery from post-mortem nail samples

Laura Jane Heathfield[1], Tia Ackermans[1,2], Aqeelah Salie[1], Tracy Bennett[1], Kate Megan Reid[1]

[1]*Biomedical Forensic Science Unit, Division of Forensic Medicine and Toxicology, Department of Pathology, Faculty of Health Sciences, University of Cape Town, South Africa*

[2]*Avans University of Applied Sciences, Lovensdijkstraat 61-63, 4818 AJ Breda, The Netherlands*

Abstract

The identification of human remains is often challenging and poses significant challenges for many forensic services worldwide. Previous studies have shown that nails could be a valid DNA source. However, these studies used varying protocols and obtained inconsistent results. The current study aimed to optimise and evaluate the recovery of DNA from nail clippings obtained from unidentified human remains for use by the state's Forensic Pathology Services. First, the input weight of nail (2 mg, 5 mg and 10 mg) and the DNA extraction incubation time (2 hr and 16 hr) were assessed using the QIAamp® DNA Investigator kit (Qiagen, Hilden) on nails from living donors (n = 5). Further optimisation using nails from cadavers (n = 6) indicated that 2 mg of nail with 16 hr incubation consistently yielded informative DNA profiles. The optimised protocol was then assessed on nail clippings from ten unidentified human remains from Salt River Mortuary, and full DNA profiles (concordant with control buccal swabs) were obtained for all cases. Lastly, the protocol was assessed on stored nail samples from an additional 45 unidentified human remains from 2018 – 2021, and full DNA profiles were obtained for 100% of nail samples. The time between death and DNA extraction did not significantly impact the DNA concentration or degradation index. These results, combined with the ease of sampling, advocate for the use of nail samples as an identification tool to help reduce the burden of unidentified human remains.

Keywords

Unidentified human remains; STR analysis; forensic DNA profiling; human identification; nail samples

Highlights

- A DNA extraction protocol was optimised for 2 mg of input nail sample.
- Full DNA profiles were generated from nail clippings in 100% of authentic forensic cases.
- Time between death and sample collection did not impact the recovery or quality of DNA.

1. Introduction

The number of unidentified bodies in forensic mortuaries was recently described as a global health crisis [1]. DNA profiling is routinely used for identifying human remains and multiple studies have shown promising results for nails as a DNA source [2–4] the identification of deceased is done by visual recognition, fingerprints, physical distinguishing marks (e.g. tattoos, scars and surgical implants. However, most previous studies have been conducted on nails samples obtained from living individuals, and the few studies on post-mortem samples had relatively small sample sizes. The amount of input nail as well as the DNA extraction protocols have varied considerably, with studies on post-mortem samples using at least 10 mg of nail, and often requiring the dissection and removal of the entire nail from the body. However, in the forensic setting, sample availability is often limited, and samples may need to be used for multiple analyses (e.g. toxicology, radioisotopes and DNA analyses). Further, there is no consensus in the literature regarding the optimal DNA extraction procedure to use on post-mortem nail samples and whether a smaller input weight of nail would yield informative forensic DNA profiles. The aim of this study was to optimise a DNA extraction protocol for small amounts of nail clipping samples and evaluate it using samples from authentic unidentified human remains obtained from the State's forensic mortuary.

2. Methods

2.1 Sample collection

For optimisation purposes, fingernail clippings were collected from all ten fingers from five living donors who had all given informed consent. Both males and females above 18 years old were included. Living donors also provided a buccal swab as a reference samples. To further optimise the method, nail clippings were collected from six cadavers. These individuals had donated their bodies to science

and provided consent during in their lifetime for their tissues to be used. The donors had been deceased between 17 and 41 days at the time of sample collection. A reference buccal swab was also obtained from cadavers.

To assess the protocol on authentic forensic cases, nail clippings and reference buccal swabs were obtained from ten unidentified deceased individuals who demised between 27 February 2021 and 24 April 2021, and who were admitted to Salt River Mortuary, Cape Town. Subsequently, stored nail clippings from a further 45 unidentified deceased individuals were available, representing individuals who demised between 1 January 2017 and 13 March 2021. The bodies from these 45 cases were no longer available for sampling of a reference swab. All samples from unidentified human remains were collected by Forensic Pathology Services (FPS) staff members in 2019 and 2021 for the purposes of identification.

Nail clippers were used to cut the free edge of all nails. The nail clippers were thoroughly cleaned in between each use with 70% ethyl alcohol and molecular biology grade water. This study received ethical approval (HREC REF: 222/2019).

2.2 Sample preparation

Nail clippings from the living donors were used for optimisation, where 2 mg, 5 mg and 10 mg sample mass were each tested with a 2 hr and 16 hr lysis incubation time. However, 2 hr incubation of 10 mg nail was not assessed as it proved insufficient time for lysis. Additional testing of 2 mg of nail clippings from the cadavers were tested with both incubation times. For unidentified human remains, 2 mg of nail was used with 16 hr incubation. Nail clippings were weighed using the AS220.R2 analytic balance scale (Radwag, Poland). To remove possible exogenous DNA, the nails were washed in 500 µl of 1X PBS (Lonza, Walkersville) and 0.5 µl Tween® 20 (Sigma-Aldrich, USA), for 10 min at room temperature with shaking at 700 rpm using the ThermoMixer F2.0 (Sigma-Aldrich, USA).

2.3 DNA extraction

DNA extractions were performed on washed nail clippings using the QIAamp® DNA Investigator Kit (Qiagen, Hilden) according to the technical manual with the following deviations: the overnight incubation at 56°C at 900 rpm shaking was done on the ThermoMixer F2.0 (Sigma-Aldrich, USA) for 2 hr or 16 hr for nails and 1 hr for swabs. The silica membrane was incubated for 3 min at 56°C before Buffer ATE was added, and after 5 min the final centrifugation was at full speed of 14000 rpm. All centrifugation steps were performed using the Centrifuge 5417C (Eppendorf, Germany). The purified DNA was diluted in 35 µl Buffer ATE for

nails and 50 µl for swabs. DNA was stored at 4 °C until further analyses. Negative controls for both sample types were processed in each batch.

2.4 DNA quantification

All samples and negative controls were quantified using the Quantifiler™ Trio DNA Quantification kit (Thermo Fisher Scientific, USA) on the 7500 Real-Time PCR System (Applied Biosystems, CA, United States) according to the manufacturer's instructions. The results were analysed with the HID Real-Time PCR Analysis Software v1.2 (Applied Biosystems, USA). The average concentration of each autosomal marker was calculated to determine the concentration of the samples. The ratio between the two autosomal amplicons gave the degradation index (DI), which was interpreted according to Vernarecci et al., 2015 [5].

2.5 DNA profiling

DNA profiling was initially carried out using the Promega PowerPlex® ESX 16 kit according to the manufacturer's handbook. For samples that initially did not yield a full DNA profile, DNA profiling was repeated using the Thermo Fisher GlobalFiler™ PCR Amplification kit. For all reactions, 0.5 ng DNA was amplified in half volume reactions on the T100 Thermal Cycler (BioRad, USA). WEN ILS500 or GeneScan™ 600 LIZ™ v2.0 was used as the internal lane standard. PowerPlex® ESX 16 amplicons were separated using the ABI 3130xl (POP7, 50 cm array) and visualised with GeneMapper v4.1 software. GlobalFiler™ amplicons were separated using the ABI 3500xl (POP4, 36 cm array) and visualised using GeneMapper® ID-X software. The analytical threshold (AT) was 66 RFU based on previous internal validations studies.

2.6 Data analysis

Shapiro-Wilk normality tests were performed. T-tests were carried out to assess any significant difference in DNA concentration and DI between input mass and incubation time for living donors and cadavers. The DNA profiles were compared to the reference profile to assess concordance. The number of successful DNA profiles was determined. A full profile required >80% of markers with peaks meeting the AT, and a partial profile were those with 20-80% of peaks that met the AT. Lastly, for data from unidentified human bodies, the Spearman rank correlation test was performed to assess if DNA concentration and DI were significantly associated with time between death and DNA extraction. GraphPad Prism (Boston, MA, USA) as used to visualise the results.

3. Results

DNA retrieved from living individuals yielded concentrations ranged between 0.0620 ng/µl and 1.9411 ng/µl, and DNA was considered only mildly degraded. No significant differences were observed between nail sample mass and incubation time for DNA concentration or DI (Supplementary Figure 1). Full DNA profiles were obtained from all living donor samples in this phase of testing, motivating to assess 2 mg of nail sample from cadavers.

DNA samples from cadavers had concentrations between 0.1130 ng/µl and 1.2685 ng/µl and were also considered mildly degraded. There were no significant differences in DNA concentration or DI from 2 mg nail clippings between each incubation time (Supplementary Figure 2). Full DNA profiles were obtained from 5 of the 6 (83%) cadaver nail samples when 2 hr incubation was used, and all 6 (100%) nail samples when 16 hr incubation was used.

DNA concentration, DI and DNA profile data from the 55 unidentified human bodies is provided in Supplementary Table 1. DNA concentration ranged from 0.1 ng/µl to 33.95 ng/µl (mean = 3.07 ng/µl) and DI ranged between 0.7 – 12.3 (mean = 1.8). Only one sample had a DI greater than 4 (indicating degraded DNA) while the majority of samples were either not degraded or mildly degraded.

Initially, 49/55 (89.09%) samples yielded full DNA profiles with the remaining 6/55 (11%) resulting in partial profiles. Upon repeating DNA profiling with the GlobalFiler™ PCR Amplification kit, the remaining 6 samples yielded full DNA profiles. Concordance was found between all nail samples and reference buccal samples.

The average time between death and DNA extraction from nails was 699 days (range: 26 - 1445 days) and did not significantly impact the DNA yield obtained. The Spearman's r value indicated a very weak positive correlation between DNA yield and the time between death and DNA extraction (Supplementary Figure 3a). There was also a very weak negative correlation between DI and the time between death and DNA extraction (Supplementary Figure 3b). Lastly, the time between death and DNA extraction did not have a significant impact on the DNA profiling success rate.

4. Discussion

In the current study, samples were obtained from bodies that had a time range of 26 – 1445 days between death and DNA extraction, and it was found that no correlation existed between time and DNA yield or DI. DNA profiling was ultimately suc-

cessful in 100% of cases, surpassing the success rate reported in other studies [2–4] the identification of deceased is done by visual recognition, fingerprints, physical distinguishing marks (e.g. tattoos, scars and surgical implants. For example, Uerlings *et al.* (2021) obtained full profiles for 80% of the samples with a post-mortem interval (PMI) between 0 – 20 weeks [2] and Inkret *et al.* (2020) obtained for 96.97% (n =32/33) full profiles on human remains with a postmortem interval up to five years [4]. This optimised DNA extraction protocol was the first to use as little as 2 mg of nail sample in DNA extraction, without compromising DNA profile success. These results support the use of nail samples as a non-invasive source for DNA profiling in a forensic post-mortem setting to help reduce the burden of unidentified human bodies.

References

[1] H. Suwalowska, J. Ali, J. Rangel de Almeida, S.A. Fonseca, L.J. Heathfield, C.A. Keyes, R. Lukande, L.J. Martin, K.M. Reid, V. Vaswani, H. Wasti, R.O. Wilson, M. Parker, P. Kingori, "The Nobodies": unidentified dead bodies—a global health crisis requiring urgent attention, Lancet Glob. Heal. 11 (2023) e1691–e1693. doi: doi.org/10.1016/S2214-109X(23)00420-5.

[2] S. Uerlings, V. Welter, B. Madea, M. Grabmüller, Comparative analysis of DNA extraction processes for DNA-based identification from putrefied bodies in forensic routine work, Forensic Sci. Int. 320 (2021) 1–6. doi: doi.org/10.1016/j.forsciint.2021.110707.

[3] A. Schlenker, K. Grimble, A. Azim, R. Owen, D. Hartman, Toenails as an alternative source material for the extraction of DNA from decomposed human remains, Forensic Sci. Int. 258 (2016) 1–10. doi:doi.org/10.1016/j.forsciint.2015.10.025.

[4] J. Inkret, E. Podov, T. Zupanc, I. Zupani, Nails as a primary sample type for molecular genetic identification of highly decomposed human remains, Int. J. Legal Med. (2020) 1629–1638. doi:https://doi.org/10.1007/s00414-020-02289-x.

[5] S. Vernarecci, E. Ottaviani, A. Agostino, E. Mei, L. Calandro, P. Montagna, Quantifiler® Trio Kit and forensic samples management: A matter of degradation, Forensic Sci. Int. Genet. 16 (2015) 77–85. doi:doi.org/10.1016/j.fsigen.2014.12.005.

Conflict of interest statement

None

Acknowledgements

A/Prof Vicky Gibbon and Mr Jacques Jacobs for their assistance with the sample collection from cadavers at the Department of Human Biology. Salt River Mortuary staff for assistance with sample collection from unidentified human remains. Yuvika Vandayar for assistance in the laboratory.

Funding

Cape Higher Education Consortium (CHEC) / Western Cape Government (WCG) joint task team research grant 2019. This work is also based on the research supported in part by the National Research Foundation of South Africa. The opinions, findings and conclusions or recommendations expressed herein are those of the author's and are not representative of the National Research Foundation

Role of the funding source

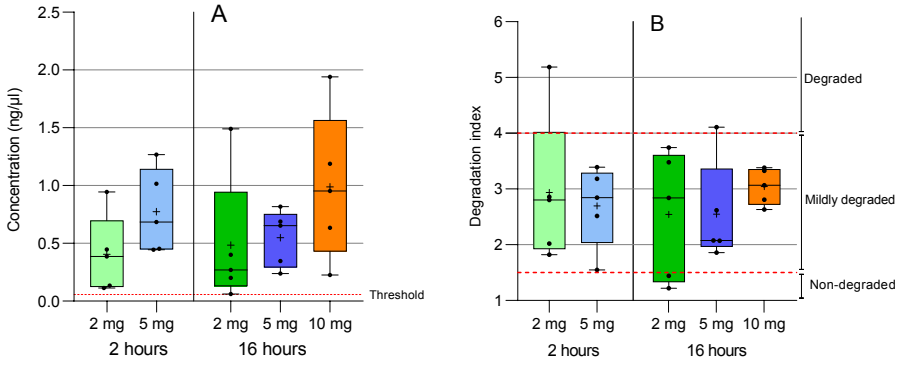
None

Supplementary Material

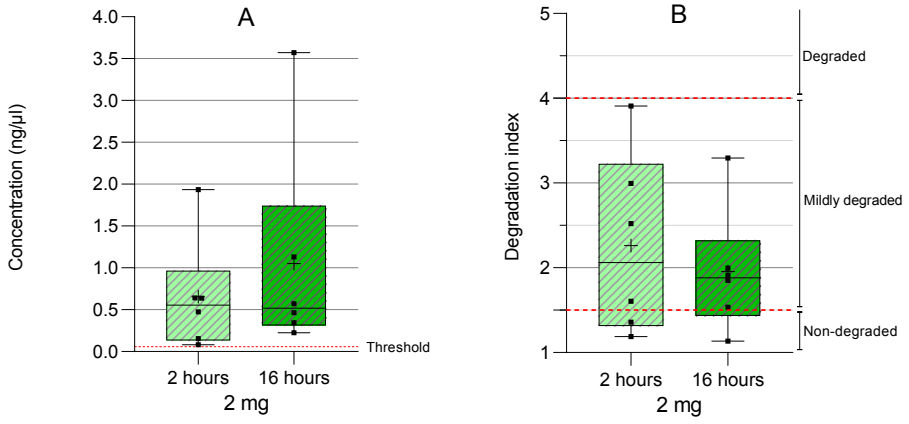
Supplementary Table 1 DNA quantification results and DNA profiling success on nails obtained from 55 unidentified human remains

Sample	IPC C_T	Degradation index	DNA concentration (ng/lL)	DNA yield (ng)	Number of successful markers /15	Number of successful markers /21
1	26.83	1.39	2.51	87.98	15	-
2	27.08	1.62	1.13	39.61	15	-
3	27.30	1.98	0.40	13.91	15	-
4	27.56	1.21	1.23	42.92	15	-
5	27.38	1.24	0.86	29.93	15	-
6	26.98	2.98	0.03	0.91	14	-
7	27.13	1.34	0.84	29.48	15	-
8	27.04	2.29	0.20	6.83	14	-
9	27.10	1.62	0.11	3.75	15	-
10	26.93	1.34	1.54	53.96	15	-
11	28.27	0.72	4.09	143.05	15	-
12	27.69	0.89	0.72	25.13	6	21
13	29.15	0.86	15.68	548.68	15	-
14	29.07	0.83	9.23	323.01	15	-
15	28.90	0.79	3.48	121.88	15	-
16	27.57	1.37	0.34	11.84	15	-
17	28.47	0.91	8.47	296.54	15	-
18	27.37	1.46	1.06	37.01	15	-
19	27.82	1.75	2.18	76.40	15	-
20	26.87	12.31	0.69	24.19	9	21
21	27.74	3.23	0.05	1.60	15	-
22	28.80	0.76	4.11	143.79	15	-
23	27.63	2.76	0.68	23.89	15	-
24	27.86	0.89	0.58	20.47	15	-
25	27.51	3.22	0.09	3.00	5	21
26	27.71	1.81	1.20	41.96	15	-
27	27.51	2.20	0.39	13.79	5	21

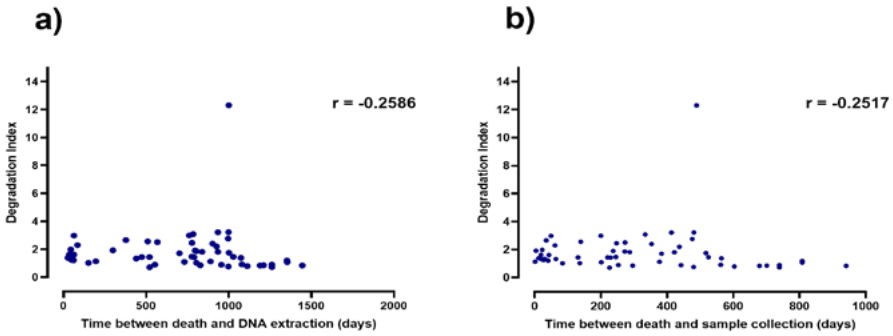
28	28.90	1.12	33.95	1188.16	15	-
29	29.04	0.86	6.81	238.42	15	-
30	27.68	1.87	0.44	15.50	15	-
31	27.69	1.47	0.25	8.84	15	-
32	27.99	1.09	0.43	14.91	15	-
33	27.93	2.51	0.04	1.34	15	-
34	28.74	0.89	4.51	158.01	15	-
35	29.92	0.71	11.13	389.51	15	-
36	27.64	1.43	1.53	53.49	15	-
37	27.68	1.43	0.53	18.54	15	-
38	28.11	2.66	3.85	134.59	15	-
39	27.72	1.92	0.08	2.82	15	-
40	29.40	0.83	9.15	320.13	15	-
41	27.98	1.19	0.01	0.36	11	21
42	27.56	1.07	0.07	2.56	15	-
43	27.59	2.40	0.14	4.79	15	-
44	27.84	1.82	0.22	7.53	14	-
45	27.93	1.03	0.15	5.29	15	-
46	27.64	1.90	0.24	8.31	15	-
47	28.03	1.43	0.21	7.35	15	-
48	27.55	2.46	0.07	2.37	10	21
49	27.38	2.98	0.04	1.37	15	-
50	27.53	1.71	0.27	9.34	15	-
51	27.57	2.56	0.03	1.12	15	-
52	28.06	1.33	0.55	19.32	15	-
53	27.74	1.13	0.10	3.42	15	-
54	27.22	3.09	0.26	9.15	15	-
55	27.92	1.02	0.22	7.57	15	-



Supplementary Figure 1: Boxplots of (A) DNA concentration and (B) DI obtained from nail clippings from living donors (n = 5) during the optimisation phase of the study. The minimum, first quartile, median, third quartile and maximum are indicated as lines on the box plot. The dots indicate individual data points and the mean is indicated with a plus(+) symbol in the boxplot.



Supplementary Figure 2: Boxplots of (A) DNA concentration and (B) DI obtained from nail clippings from cadavers (n = 6) during the optimisation phase of the study. The minimum, first quartile, median, third quartile and maximum are indicated as lines on the box plot. The dots indicate individual data points and the mean is indicated with a plus(+) symbol in the boxplot.



Supplementary Figure 3: Scatterplots of the time between death and DNA extraction plotted against (A) DNA yield and (B) DI obtained from nail clippings from unidentified human remains (n = 55).

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 127-133
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Exploring the Effects of Temperature-Dependent Mitochondrial DNA Sequencing from Incinerated Tooth Samples: A Preliminary Study

Kangana Aggarwal[1], Nidhi Sharma[1] Kulbhushan Prasad[1], Sudhir Kumar Gupta[1], B.K. Mohapatra[2], Kamal Chauhan[2], Harpreet Singh[3], Chittaranjan Behera[1*]

[1] *All India Institute of Medical Sciences, New Delhi, India.*
[2] *Central Forensic Science Laboratory, DFSS, MHA, New Delhi, India.*
[3] *Indian Council of Medical Research, New Delhi, India.*

** Corresponding author at: All India Institute of Medical Sciences, New Delhi, India-110029, drchitta75@rediffmail.com*

Abstract

Teeth serve as crucial evidence for human identification in incidents like explosions, mass fatalities, and fires, where the body is severely compromised and antemortem dental records may be unavailable. Their DNA-rich structure makes them ideal for analysis, though nuclear DNA amplification often becomes difficult due to degradation. In such cases, mitochondrial DNA (mtDNA) is useful as it is more resistant to degradation and present in high copies, aiding maternal identification. This study aimed to perform mtDNA typing from incinerated tooth samples where nuclear DNA profiling was partial or unreliable. Thirty molar teeth were incinerated at 100-300°C for 15 minutes, and DNA was extracted using organic extraction method. DNA quantity and quality were assessed with a Nanodrop 1000 and agarose gel. STR profiling and mtDNA amplification were conducted, followed by Sanger sequencing. It Was observed that complete STR profiles were obtained at 100°C and 200°C, while partial profiles appeared at 300°C, indicating degraded nuclear DNA. However, mtDNA amplification succeeded across all temperatures, demonstrating that mtDNA can be recovered from incinerated samples even when nuclear DNA is insufficient.

Keywords

mtDNA, incinerated, tooth, haplotypes.

Introduction

Teeth are the hardest and most resistant to putrefaction even after many decades. They are important evidence for genetic identification as its structural components are highly rich in DNA which makes it suitable for DNA analysis. In cases like explosions, mass death incidents, suicide attempts, accidents by car or aircraft, house fires or arson, the integrity of the body gets severely disrupted and it is also not necessary that antemortem dental records will be available. Therefore, the identification of the person through visual inspection becomes unfeasible. So, for the identification of the deceased from the remains of the body genetic analysis is performed. In such cases, the amplification of genetic markers of genomic DNA becomes difficult and unfeasible due to degradation of nucleic acids. In such cases, mitochondrial DNA is a useful alternative as it is used as supporting evidence for identification of maternal relationship. Its characteristics like resistance to degradation and high number of copies per cell helps in obtaining a high yield. Some published literature suggests that recovery of DNA is feasible at high temperature as high as 300ºC and even mtDNA can be detected. But to be a certain, mtDNA typing is necessary to help in forensic investigations [1-3]. Therefore, the aim of the study is to perform mitochondrial DNA typing from incinerated tooth samples where results obtained from nuclear DNA profiling is either partial or non-reliable.

Material studied, methods, techniques

Sample collection and preparation

The study was approved by Ethical Sub Committee with ref no. IECPG-127/24.02.2022 by All India Institute of Medical Sciences, New Delhi, India. A total of 30 molar tooth samples, (n=30), where 3 tooth samples were collected from each subject by obtaining informed consent from their immediate family members and relatives before the autopsies at the Department of Forensic Medicine and Toxicology, AIIMS, New Delhi. These samples were exposed to high temperatures from 100-300ºC, respectively, by using muffle furnace for 15mins each and were stored at room temperature.

Pre-processing of samples

After incinerating all the tooth samples, each sample was pulverised into fine powder by using Tissue Lyser (Qiagen). 0.5mg of each sample was taken and

transferred into a fresh tube of 1.5mL. The samples were decalcified by using 0.5M EDTA and were incubated at 56ºC at 800rpm for 12hrs till the completion of 72 hrs.

DNA extraction and Isolation

To the samples, 500µL of lysis buffer (2% CTAB (20 g), 2.56 NaCl, 20mM EDTA, 0.1M Tris-HCl (pH to 8.0) was added along with the 50µL of 20mg/ml proteinase k. Proteinase k was added in intervals of 24hrs. The samples were kept in dry bath for incubation at 56ºC for 48 hrs at 10,000 rpm. The organic extraction was performed where phenol, chloroform, and isoamyl alcohol were used in the ratio of 25:24:1. The samples were precipitated by using 3M sodium acetate (pH 5.5) and equal volumes of isopropyl alcohol, followed by washing with 70% ethanol. The DNA samples were eluted in 20µL of 1X TE buffer. All DNA samples were stored at -20ºC till further processing.

Qualitative and Quantitative Analysis

DNA quality was estimated from a 1% agarose gel. The absolute yield of extracted DNA was quantified by using a Nanodrop 1000 spectrophotometer (Thermo Fischer). The quantity of DNA samples was between 110-600ng/µL, and absorbance quotient (Q) was between 1.7-2 which were considered pure and further processed for PCR.

Multiplex PCR for STR profiling

STR profiling was performed by using the AmpFLSTR™ Identifiler™ Plus PCR Amplification kit (Applied Biosystems, USA) according the manufacturer's protocol. DNA genotyping was carried out on a 3500xL Genetic Analyzer (Applied Biosystems, USA) using POP 4 and analyses of the generated autosomal STR profiles were performed using Gene Mapper HID Software Version 1.5

Mitochondrial DNA Sequencing

The same samples were amplified using 10 primer pairs for the control region of mitochondrial DNA from published literature [4]. The amplicons and 50 bps DNA marker (SMOBIO) were visualised on a 2% agarose gel to estimate the intensity of the amplification as shown in Fig. 1.

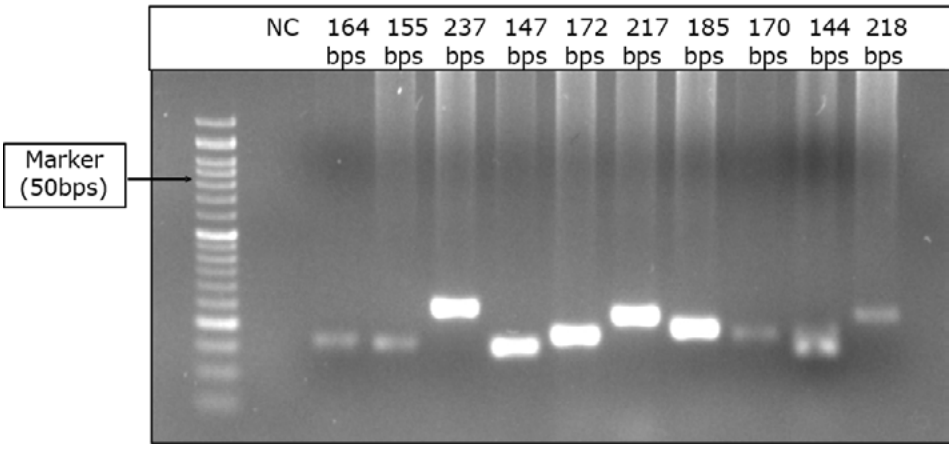


Fig 1: 2% agarose gel showing amplicons of tooth samples

PCR products were cleaned by using ExoSAP-IT (Thermo Fisher, USA) and sequenced using standardized protocols of Sanger Sequencing using Big Dye Terminator sequencing reagents (Version 3.1, Applied Biosystems). Sequencing reaction products were purified from residual dye terminators using Big Dye Xterminator purification kit, according to the manufacturer's manual. DNA sequencing was carried out on a 3500xL Genetic Analyzer (Applied Biosystems) and sequences were analysed using Sequencing Analysis Version 6.0 (Applied Biosystems). Further, haplotypes were determined from the sequences generated by Seqscape version 3.0. (Applied Biosystems) which aligned the sample sequences with the reference sequence, rCRS. The haplogroups were obtained for all the samples from EMPOP database [5].

Results

All 30 tooth samples (n=30) were extracted successfully which were incinerated at high temperatures at 15mins. Samples incinerated at 100ºC has the estimated quantity of DNA of 5.2-6.10 ng/µL with the average quantity 5.3ng/µL with which STR profiling was successful. Similarly, the quantity of DNA in samples incinerated at 200ºC was 3.10-4.50 ng/µL with an average quantity was 3.30ng/µL indicating a slight decrease in DNA quantity. Despites the low quantity, a complete STR profile was observed. But in samples incinerated at 300ºC, DNA quantity observed was 0.89-1.10 ng/µL. Due to reduced quantity of DNA, there was a slight decrease in

the number of amplified loci indicating further degradation or loss of DNA quality over high temperature showing significant dropouts with no detectable alleles in most samples.

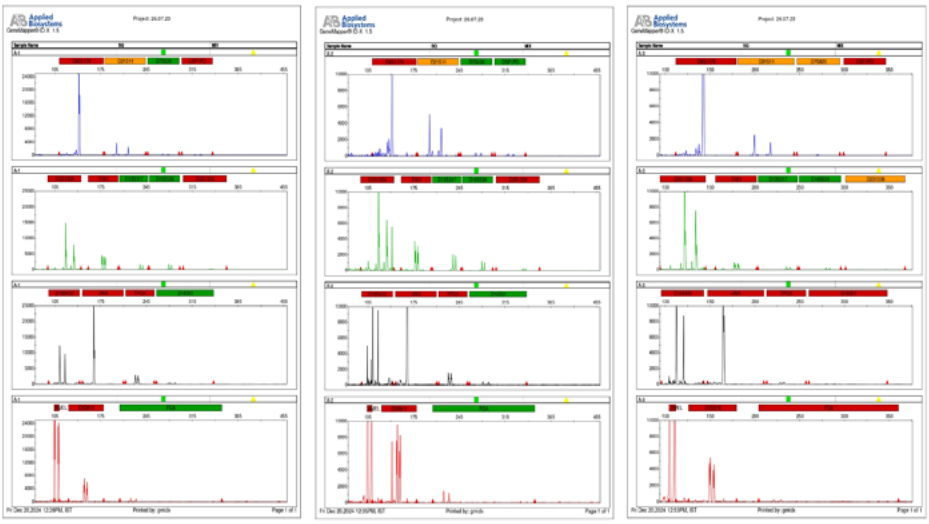


Fig 2: STR profile at 100ºC, 200ºC and 300ºC from the same subject

Mitochondrial DNA typing was performed in all the samples to assess the utility of mtDNA analysis in samples where STR profiling was unsuccessful. All sequences showed the acceptable quality of sequences with appropriate base calling and peak height.

All samples (n=30) were successfully typed for mtDNA, revealing different haplogroups, including A11, N70, L3f1a1, M40a1, and N5. These haplogroups were identified based on the presence of specific mutations by using EMPOP database. A detailed analysis of haplotypes showed that each sample had a distinct set of mutations, with common mutations such as 73G, 263G, 310C, 315.1C, 16223T, and 16519C.

The results demonstrate that while STR profiling was largely unsuccessful due to the degradation of nuclear DNA, mtDNA typing provided reliable and consistent results. This suggests that mtDNA analysis is a more robust method for genetic profiling in cases where samples are recovered in cases of mass disaster and explosions, which often leads to significant DNA degradation.

Discussion

The results of this study align with previous research, reinforcing the utility of mitochondrial DNA (mtDNA) in forensic identification, especially in challenging and degrading conditions like high temperatures. Data from published literature demonstrates that mtDNA is more resilient to thermal degradation than nuclear DNA, making it a reliable alternative when short tandem repeat (STR) profiling is compromised as DNA molecules are fragmented [1]. Similarly, in another study highlighted the importance of mtDNA analysis in forensic cases involving burnt remains, as it remains intact and can be amplified even in severely degraded samples [3].

In this study, STR profiles were complete in samples incinerated at 100°C and 200°C, consistent with findings by [2], who observed that nuclear DNA starts degrading significantly at higher temperatures. However, samples incinerated at 300°C exhibited partial profiles, reducing the efficacy of nuclear DNA typing as increased thermal exposure increases. Despite this, mtDNA amplification and sequencing were successful across all temperature conditions, underscoring the robustness of mtDNA in extreme environments.

Conclusion

This study concludes that mitochondrial DNA can be recovered and sequenced from incinerated samples exposed to high temperatures, especially in cases where results obtained from STR profiling are either partial or unreliable for identification purposes.

Acknowledgments

This project was funded by Indian Council of Medical Research, New Delhi, India under project code I-1063.

Conflict of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Devaraju, R. Gantala, M. Ambati, A. Vemula, J. Kubbi, S. Gotoor, DNA detection in tooth exposed to different temperatures: An in vitro study, J Indian Acad Oral Med Radiol 26 (2014) 393. https://doi.org/10.4103/0972-1363.155681.

[2] N.A. Mahat, K. Dhevaseelan, N.A. Sharif, H.O. Khan, A.R. Azman, E. Layang, S.A. Ismail, Recovery of human DNA from canine teeth exposed to direct heating of 300 °C at varying durations for forensic identification, in: Bali, Indonesia, 2019: p. 020004. https://doi.org/10.1063/1.5125508.

[3] E. Yukseloglu, K. Dastan, F. Yonar, G. Rayimoglu, O. Karatas, D. Islek, M. Dogan, The comparison of DNA extraction techniques in human bone and tooth samples exposed to high heat, Med-Science (2019) 489. https://doi.org/10.5455/medscience.2019.08.9051.

[4] C. Eichmann, W. Parson, "Mitominis": multiplex PCR analysis of reduced size amplicons for compound sequence analysis of the entire mtDNA control region in highly degraded samples, Int J Legal Med 122 (2008) 385–388. https://doi.org/10.1007/s00414-008-0227-5.

[5] W. Parson, A. Dür, EMPOP—A forensic mtDNA database, Forensic Science International: Genetics 1 (2007) 88–92. https://doi.org/10.1016/j.fsigen.2007.01.018.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 135-141
DOI: https://dx.doi.org/10.15304/cc.2025.1869

AI-based, Automated Detection of Spermatozoa in Standard Forensic Specimens

Christian Schunck[1*], Konstantin Ehinger[1]

[1]*MetaSystems Hard & Software GmbH, Altlussheim, Germany.*

** Corresponding author at: tel. +49 6205 39610, cschunck@metasystems.de*

Abstract

1. Objectives

In the field of image analysis, artificial intelligence, particularly using deep neural networks (DNNs), has shown significant effectiveness. These networks, trained on large collections of categorized image data, achieve high accuracy, for example, in detecting and identifying objects within digital images of microscopic samples.

This advancement holds particular importance in forensic investigations related to suspected sexual offenses, where analyzing Rape Kits and other evidence for the presence of sperm is a standard procedure. Traditionally, forensic labs rely on manual microscopy, a process that is both time-consuming and labor-intensive, requiring the expertise of specially trained personnel. The manual approach contributes significantly to backlogs in case processing due to its demands on time and resources.

2. Material and Methods

*T*o address these issues, we have developed an automated workflow leveraging two DNNs trained for sperm detection and classification, aimed at streamlining the analysis of trace evidence. These networks were trained on forensic specimens prepared using common staining techniques, including “Christmas Tree” stain, Baecchi stain, and H&E, through supervised learning.

3. Results and Conclusions

Internal examinations using Baecchi-stained samples from a standard forensic laboratory demonstrated a detection sensitivity of 98.7% and a classification accuracy of 98.4%. A direct comparison of manual evaluation methods and DNN-based automatic microscopy on 80 specimens in a European forensic laboratory showed that 10 cases, previously identified as negative through manual microscopy, were positive.

These results indicate that our AI-enhanced method for identifying sperm on microscopic slides presents a viable option for decreasing the time and resources required by conventional microscopy. This, in turn, could help alleviate the backlog problem faced by forensic laboratories.

Keywords
Sperm Detection, Digital Microscopy, Automation.

Introduction

Manual microscopy, long regarded as the gold standard for forensic sperm detection, presents several notable shortcomings that can hinder the efficiency of forensic investigations. This method requires significant time and specialized expertise, as trained forensic analysts must visually inspect samples under a microscope to identify spermatozoa. The process is labor-intensive and prone to human error due to factors such as fatigue, subjectivity, and variability in training among personnel.

The reliance on manual microscopy in forensic laboratories is likely one of the key factors contributing to delays in case resolution, particularly in the analysis of sexual offense cases. As case volumes increase, so does the burden on forensic experts, exacerbating delays in the criminal justice system. The extensive time needed to analyze slides manually means that each case adds to the accumulation of unsolved cases, creating backlogs that can take months or even years to resolve [1].

Automation of imaging offers a potential solution to this challenge. Automated systems are capable of processing large volumes of samples with speed and precision, minimizing human error while dramatically reducing the time required for analysis. In the field of image analysis, the application of artificial intelligence, particularly using deep neural networks (DNNs), has demonstrated remarkable efficacy [2].

We employed a commercially available microscope-based imaging software, Metafer (MetaSystems, Germany), which supports the integration of DNNs into its scanning workflow [see Fig. 1]. Leveraging this platform, we developed an automated process incorporating two DNNs specifically trained for sperm detection and classification, aimed at enhancing trace evidence analysis. These networks were trained through supervised learning using forensic samples prepared with common staining techniques, including Nuclear Fast Red/picroindigocarmine ("Christmas Tree"), acid fuchsine/methylene blue ("Baecchi"), and hematoxylin/eosin (H&E) stains. Our workflow features two distinct DNNs: one for object detection and the other for object classification. During the scanning process, Metafer applies both networks to generate object images and outputs a classification score, indicating the level of similarity with true sperm cells from the training data set.

Fig. 1: Installation of Metafer for forensic sperm detection. Left of the microscope, the robotic SlideFeeder x80 device allows for enhancing the capacity of the scanning installation to up to 800 slides per run.

Material studied, methods, techniques

Scanning Software and Integrated Hardware

Automated imaging was conducted using the Metafer slide-scanning software (MetaSystems, Altlussheim, Germany). This commercially available solution integrates a motorized microscope (AxioImager Z2, Carl Zeiss, Jena, Germany) equipped with transmitted light LED illumination, a motorized X/Y scanning stage (Maerzhaeuser, Wetzlar, Germany), an ultra-high-resolution color camera (CoolCube 4Pc; MetaSystems, Altlussheim, Germany), and a Windows™-based PC (DELL, Langen, Germany) running the Metafer software.

The software directly controls various hardware components, such as the microscope's focusing motor, objective turret, and motorized stage. Slide-scanning and automated image assessment parameters were configured based on the experiment's requirements by selecting the appropriate settings file (classifiers) in the Metafer software. These classifiers define key parameters for image acquisition, such as the number of captured fields per sample, magnification, and more. Additionally, classifiers enable the specification of the DNNs used for object detection and classification.

To generate DNN training images and for the imaging process itself, we selected the EC Plan-Neofluar 20x/0.50 M27 microscope objective (Carl Zeiss, Jena, Germany) as the most suitable option.

The capacity of the scanning installation ranges from 8 regular microscope slides per run to up to 800 slides per run, if an optional SlideFeeder x80 for automated exchange of slide frames is added to the installation [see Fig. 1].

DNN Development

We employed supervised learning to train the deep neural networks (DNNs) used in the sperm detection workflow. In this process, DNNs learn from example images that have been labeled by forensic experts. During training, the DNN optimizes its predictions by comparing its predicted labels to the actual labels provided by the expert (the "ground truth"). This supervised learning approach allows for continuous monitoring of the DNN's training progress. Upon completion of training, the DNN was tested on a set of previously unseen images (approximately 5% of the dataset), which had been set aside before training.

A dedicated DNN, specifically trained for spermatozoa detection, identifies potential sperm cells. This network was designed with a high sensitivity to minimize the risk of missing spermatozoa. As a result, it also detects several background objects that resemble spermatozoa. These objects appear in the image gallery but typically receive a low score from a subsequent DNN, which performs further classification. The detection network was trained on approximately 20,000 objects derived from over 2,000 fields of view (FOV).

A second DNN, responsible for classifying spermatozoa among the objects detected by the first DNN, was trained on nearly 80,000 objects from around 350 original slides stained with "Christmas Tree" stain (approximately 63%), "Baecchi" stain (28%), and H&E stain (9%). This classification DNN evaluates the detected objects and categorizes them into two groups: (1) potentially sperm-cell-negative and (2) potentially sperm-cell-positive. It assigns a score to each object, indicating the similarity to the training data of each of the two categories, and organizes the image gallery accordingly. All detected objects, even those with low scores, are included in the results gallery for expert review via the Metafer software.

Both DNNs are seamlessly integrated into the imaging process, with evaluation occurring in real time as the software scans the slides.

Results

The proposed workflow consists of the automated digitization of standard forensic samples for sperm detection with the help of the slide scanning software installation. During the scanning procedure, a DNN for object detection identifies spermatozoa-like objects and marks their positions. A second DNN, intended to classify the objects, records the similarity of the object to the training data of the target classes, namely (1) potentially sperm-cell-negative and (b) potentially sperm-cell-positive.

An image of each object is taken by the scanning software, and its position on the slide is recorded. After the scan, a gallery of all objects is presented and sorted by the classification score for potentially sperm-cell-positive objects.

The performance of our DNN-based sperm detection workflow was evaluated using a dataset of Baecchi-stained samples collected from a standard forensic laboratory. The DNNs were trained and tested under supervised learning conditions (see above), leading to the following findings:

- **Detection Accuracy and Sensitivity:** In our internal tests, the sperm detection DNN achieved a sensitivity of **98.7%** and an accuracy of **13.6%**. The high sensitivity underscores the model's effectiveness in identifying sperm cells in stained microscopic images, while the relatively low accuracy is designed to minimize the occurrence of false-negative results.
- **Classification Accuracy and Sensitivity:** The classification DNN, which is tasked with categorizing the detected objects, achieved an accuracy of **91.9%** and a sensitivity of **98.4%** in our internal tests. This level of accuracy demonstrates the DNN's capability to distinguish between potentially sperm-positive and sperm-negative objects effectively.

Discussion

Microscope-based detection of spermatozoa in forensic samples, such as those from Rape Kits or other trace evidence, remains a critical component in the prosecution of sexual assault cases. However, the time-consuming and labor-intensive nature of this procedure, coupled with the increasing backlog in forensic laboratories worldwide, has led to numerous attempts to streamline this step. Alternative methods such as PCR [1], specialized fluorescence markers [2], and even cell phone-based techniques [3] have been explored.

Our objective was to accelerate the sperm detection process in forensic laboratories while maintaining the established method of microscopy. To achieve this, we developed an automated imaging workflow for the detection and classification of spermatozoa using a microscope. This workflow was deliberately designed around standard forensic samples stained with widely accepted techniques, thereby eliminating the need for specialized sample preparation. The system utilizes commercially available slide scanning software and incorporates Deep Neural Networks (DNNs) for both detection and classification of spermatozoa.

The workflow integrates two DNNs with automated microscope imaging software, enhancing the efficiency of the microscopy process in forensic labs. It generates a gallery of detected and classified objects, ranked by their likelihood of being sperm cells, thus providing forensic experts with a comprehensive, machine-generated output for rapid on-screen verification. Additionally, the system preserves the captured image fields, facilitating enhanced documentation.

Conclusion

While the performance of our DNN-based workflow is influenced by the quality of the samples, with validation remaining the responsibility of individual laboratories, our internal evaluations using routine "Baecchi"-stained forensic samples have demonstrated high sensitivity and accuracy in DNN-based sperm detection. These results are further supported by reports from laboratories that have tested or adopted the workflow into their standard practices. For example, a European laboratory conducted a direct comparison between manual sperm detection methods and our DNN-based workflow on 80 slides from their routine workload. The DNN workflow successfully identified 9 samples that had been mistakenly classified as negative but actually contained spermatozoa (data unpublished). This aligns with findings from other institutes, as presented in digital posters at the ISFG 2024 conference in Santiago de Compostela, Spain, in September 2024.

Conflict of interest statement

The authors are employees of MetaSystems Hard & Software GmbH, Altlussheim, Germany.

References

[1] S. Riman, C.H. Shek, V. Clermont Beaudoin, *et al*., Confirmatory detection of sperm and semen *Via* proximity ligation real-time PCR. Forensic Science International: Genetics Suppl. Series, 5 (2015) e109-e111.

[2] J. Skolnick, W. Jones, K. Alasoo, *et al*., Computational biology: deep learning. Emerg Top Life Sci 14 (2017) 257–274.

[3] A. De Moors, T. Georgalis, G. Armstrong, *et al*., Sperm Hy-Liter™: An effective tool for the detection of spermatozoa in sexual assault exhibits, Forensic Science International: Genetics, 7 (2013) 367-379,

[4] S. Deshmukh, F. Inci, M.G. Karaaslan, *et al*., A confirmatory test for sperm in sexual assault samples using a microfluidic-integrated cell phone imaging system, Forensic Science International: Genetics, 48 (2020).

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 143-148
DOI: https://dx.doi.org/10.15304/cc.2025.1869

A comparison of likelihood ratios obtained from EuroForMix and STRmix™ for two and three contributor mixture profiles

Manuela Saccente[1,2,*], Alessandra Caglià[1], Alessandra La Rosa[1], Stefano D'Amelio[2]

[1] *Forensic Science Police Service, Forensic Science Police and Cyber Security Central Directorate, Ministry of Interior, Rome, Italy*

[2] *Department of Public Health and Infectious Disease, Sapienza University of Rome, Italy*

** Corresponding author at: Forensic Science Police Service - via Tuscolana 1548, 00173 Rome, Italy*
E-mail: manuela.saccente@gmail.com, Phone: +393286894939

Abstract

DNA mixtures are routinely used to identify donors of biological samples obtained from crime scenes. If one or more persons of interest are available in a case, the DNA profile obtained from the crime scene evidence can be interpreted using probabilistic genotyping software to calculate the weight of evidence (WoE).

The differences in likelihood ratios (LR) between the probabilistic genotyping software EuroForMix [1] and STRmix™ [2] were examined on 50 two- and three- contributor mixtures, selected from the PROVEDIt dataset [3].

In most cases the two SW are aligned and the difference between the two LR values falls within two orders of magnitude.

In addition, a relative of the Person of Interest (POI) was studied as an alternative donor in the denominator proposition to highlight how the two SW evaluate the effect of kinship on the relative LR values.

Keywords

Probabilistic genotyping, DNA mixtures, EuroForMix, STRmix™

Introduction

The use of probabilistic genotyping (PGS) software to calculate the likelihood ratio (LR) has become widespread for studying mixed genetic profiles of variable and increasing complexity.

At this point it is undisputed that continuous software is the best performing method for studying mixtures, especially those in which the ratio of contributors is unbalanced, even if there is no single or best mathematical model, nor consequently an absolutely "correct" value of the LR relating to a given mixture.

In any case, it is important that the value of the calculated LR is solid and reliable, especially if it is reported in court.

The aim of the present work was to compare the LR values obtained with two different types of continous software (SW), EuroForMix (EFM) (v. 3.3.4 – 4.0.8) and STRmix™ (v. 2.9.1 - 2.11), which use different mathematical models, as suggested by the report of the President's Council of Advisors on Science and technology, PCAST [4]. In this report, PCAST argued that the comparison should be done by independent groups (i.e., not the software developers).

We believe that our study goes exactly in this direction.

Material studied, methods, techniques

We selected 50 mixtures from the PROVEDIt dataset, 30 mixtures with 2 contributors (2 PM) and 20 with 3 contributors (3 PM), with various mixture ratios (MR) (1:1, 1:2, 1:4, 1:9, 1:1:1, 1:2:2, 1:9:1 and 1:9:9), degradation (with Quality Index varying from LAND, i.e. Large Autosomal Not Detected, to 10.9), and DNA input (from 0.030 ng/ul to 0.750 ng/ul).

We calculated the LR for each contributor for 4 pairs of hypotheses [of the H1/H2 type]: the LR_{base} with the base hypothesis,

LR_{base} = H1/ H2 = (POI+U)/=(U+U) for 2PM;
or LR_{base}= H1/ H2 = (POI+U)/(U+U+U) for 3PM;

and three LRs_{Fam}, replacing an unknown person in H2 with a relative of the POI (father/son or sibilings or cousin):

The calculations were performed with different software versions because they were executed at different times (EFM v. 3.3.4 – 4.0.8 and STRmix™ v. 2.9.1 - 2.11).

The comparison was made between the Log10LR(MLE) values of EFM and the Log10LR-Lower HPD values of STRmix™.

The allele frequencies of the Italian population and the Fst-correction value set to 0.01 were used. The other setting parameters have been left at their default value. Dye-specific analytical thresholds (AT) validated within our laboratory were

applied to the raw data (lower threshold called "LT-DNA"). A fixed NOC was used since the ground truth of the mixtures was known. For three mixtures (points 13 and 14, 29 and 30 and 43 of Fig. 1) the calculation was performed applying a higher AT threshold (called "cDNA"), for technical reasons (essentially for the high number of extra allele peaks that appeared applying the LT-DNA threshold).

The Log10(LR) values obtained with the two SW were compared with each other and then also compared to three parameters: the number of drop-out, the DNA input, and degradation, in order to study and highlight the possible impact of these on the different calculation methods.

Results

From the analysis of the obtained results, it was possible to highlight a good agreement in the majority of the analysed mixtures, especially in the case of 3PM, with a maximum total average variation of the $log10LR_{base}$ ($\Delta LR_{average}$) of 0.97, compared to the 2PM, with $\Delta LR_{average}$ of 1.77 (see Tables 1).

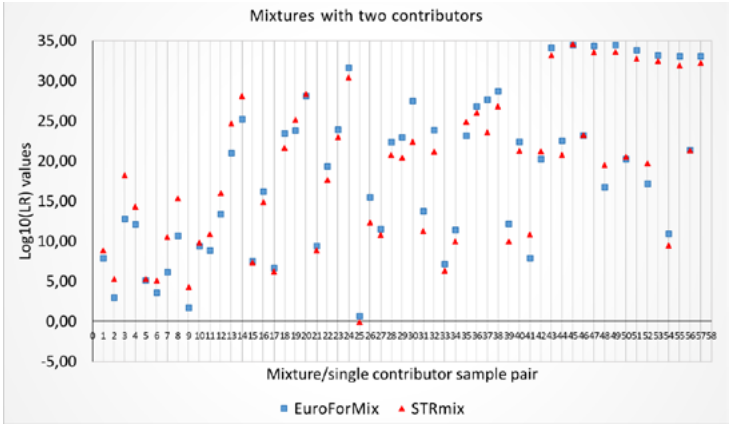
The 2PM 1:1 are the mixtures with the greatest variance in the log10LRbase, with ΔLR_{base} equal to 2.43 (see Tables 1).

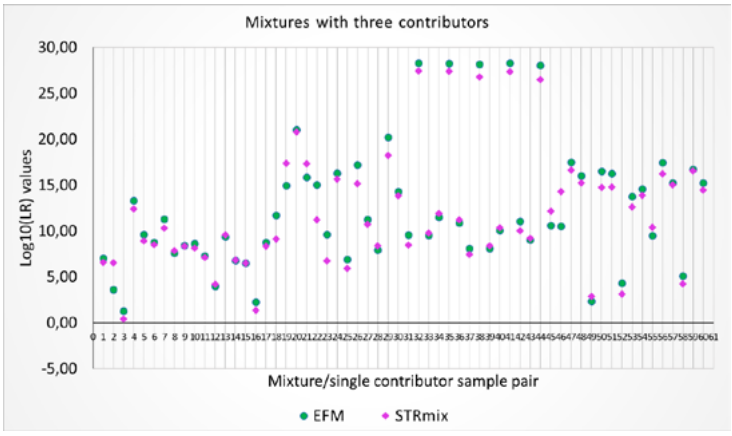
Table 1: Values of the average variation of log10LR (ΔLR) for the 4 groups of 2PM mixtures (with MR 1:1, 1:2, 1:4 and 1:9) and 4 groups of 3PM mixtures (with MR 1:1:1, 1:2:2, 1:9:1 and 1:9:9) and total average variation in log10LR ($\Delta LR_{average}$) for the four hypotheses formulated for the calculation of LR (LR_{base}, LR_{cousin}, $LR_{parents}$ and $LR_{siblings}$).

	ΔLRBase	**ΔLRcousin**	**ΔLRparents**	**ΔLRsiblings**
2PM - MR 1:1	2,43	2,36	2,02	1,62
2PM - MR 1:2	1,40	1,05	0,67	0,44
2PM - MR 1:4	2,20	1,57	1,44	0,81
2PM - MR 1:9	1,05	0,86	0,76	0,45
$\Delta LR_{average}$	**1,77**	**1,46**	**1,22**	**0,83**
3PM - MR 1:1:1	0,57	0,46	0,35	0,34
3PM - MR 1:2:2	1,46	1,17	0,88	0,86
3PM - MR 1:9:1	0,78	0,43	0,61	0,25
3PM - MR 1:9:9	1,08	1,01	0,87	0,75
$\Delta LR_{average}$	**0,97**	**0,77**	**0,68**	**0,55**

It was observed that both 2PM and 3PM ΔLR values decreased with the increase in genetic similarity between POI and the considered relative (genetically speaking siblings are more similar than cousins) (see Tables 1), suggesting a greater alignment of the performance of the two types of software in the calculation of this familiar LR.

The Log10(LR) values relating to the basic hypothesis for the 2PM and the 3PM, obtained with the EFM and STRmix SW, are shown in the two plots (see Figs. 1 and 2). Similar trends were obtained for the other three hypotheses (LRs_{FAM}).





Figs. 1-2: LR_{base} values (log10 scale) computed for the same mixture/single contributor sample pairs, with two (left plot-Fig.1) and three (right plot-Fig.2) contributors in EuroForMix (blue square and green circle) and STRmix™ (red triangle and fuchsia diamond).

The STRmix™ SW returns a total LR equal to 0 in cases where one or more specific locus LR are equal to zero (for three mixtures in this study with the application of the LT-DNA threshold– data not shown), due to the exclusion of the genotypic combination request. On the contrary, the EFM SW calculates a weight of evidence for all the various genotypic combinations, even those that are very unlikely, always returning a total LR value.

The LR values obtained with both types of SW were compared with each other, also in relation to the parameters number of drop-out, DNA input and degradation, in order to study and highlight the possible impact of these parameters on the different calculation methods. From this study it emerged that while for the degradation and number of drop-out parameters there is a certain inverse proportionality with respect to the LR values, for the DNA input quantity a trend is not immediately evident, instead (data not shown).

Discussion

Our study was performed to understand the variability of the LR calculation using the EFM and STRmix™ software. The comparison of the LR values obtained from the two different continuous PG models highlights that in most cases the two SW are aligned, especially in the case of the 3PM, and the difference between the two LR values falls within two orders of magnitude.

In EFM, some calculations carried out with the LT-DNA threshold returned an LR more similar to that obtained with STRmix™, but not reportable in court because one or both Hp and Hd Model Validations failed. On the other hand, the respective values obtained with EFM using the cDNA threshold, with successful Hp and Hd Model Validation, were more divergent than those obtained with STRmix™. This is probably due to the appearance of a high number of drop-out, and to the disappearance of most of the stutter peaks, to which STRmix seems to be more sensitive, especially in the case of the minor contributor.

In our opinion, the differences in the high LR values returned by the two SW may not cause concern to the expert geneticist, whereas when the LR values are relatively low (e.g. lower than 10^6 or 10^4), much more caution and a more in-depth investigation should be reserved.

Currently, the most divergent results (4/50 mixtures) are the subject of further studies, in order to better identify the possible reasons for these differences. The comparison study described here is also being extended to mixtures from real cases.

Conflict of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Ø. Bleka, G. Storvik, P. Gill, EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts, Forensic Sci. Int. Genet. 21 (2016) 35-44.

[2] D.A. Taylor, J.-A. Bright, J. S. Buckleton, The interpretation of single source and mixed DNA profiles, Forensic Science International: Genetics. 7(5) (2013) 516-528.

[3] Lauren E. Alfonse, Amanda D. Garrett, Desmond S. Lunb et al, A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt, Forensic Science International: Genetics. 32 (2018) 62-70.

Web references

[4] President's Council of Advisors on Science and Technology. Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods – 2016. Available from: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ ostp/PCAST/pcast_forensic_science_report_final.pdf. Accessed 27 Apr 2017.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 149-154
DOI: https://dx.doi.org/10.15304/cc.2025.1869

KongohPlus: An R-shiny package for probabilistic genotyping software based on an MCMC approach

Sho Manabe[1*], Masaki Hashiyada[1], Atsushi Akane[1]

[1] *Department of Legal Medicine, Kansai Medical University, Hirakata, Japan.*

** Corresponding author at: Department of Legal Medicine, Kansai Medical University, 2-5-1, Shin-machi, Hirakata, Osaka, 573-1010, Japan,*
Email: manabesh@hirakata.kmu.ac.jp · Tel.: +81 72 804 2412; Fax: +81 72 804 2419

Abstract

Probabilistic genotyping software based on the fully-continuous model is effective for interpreting crime stain profiles. In this study, we developed a new probabilistic genotyping software "KongohPlus" based on a Markov chain Monte Carlo (MCMC) approach. KongohPlus is an R-shiny package with interactive graphical user interface application. The MCMC models are written using the Stan language, which adopts the Hamiltonian Monte Carlo algorithm as the MCMC method. The parameters of the MCMC are mixture proportions per contributor, degradation parameters per contributor, total peak height that reflects the total DNA amount, and locus-specific amplification efficiencies. The expected allele-specific stutter ratios for back stutters, forward stutters, double-back stutters, and minus 2-nt stutters are determined based on the user-prepared experimental data. All possible genotype combinations are considered in each MCMC step; therefore, the number of warm-up steps and sampling steps are relatively small. The Rhat values (also known as the Gelman-Rubin diagnostics) of the likelihood and each parameter are used for convergence diagnostics.

The software KongohPlus was run as a pre-trial using some 2–4 person mixed DNA profiles typed by the GlobalFiler™ PCR Amplification Kit (Thermo Fisher Scientific). In the conditions that the number of warm-up steps is 1,200, the number of sampling steps is 300, and the number of chains is 4, the convergence diagnostics have been achieved (i.e., all Rhat values were smaller than 1.1). The computational time was approximate 20 sec in 2-person mixtures, approximate 10 min in 3-person mixtures, and approximate 10 hours in 4-person mixtures. The computational time was expected to be significantly reduced if rare stutter types (e.g., double-back stutters) were not considered. We are going to perform the developmental validation of KongohPlus preparing the experimental profiles typed by the GlobalFiler, which is generally used for actual caseworks in Japan.

Keywords

Probabilistic genotyping software, continuous model, short tandem repeat, DNA mixture, GlobalFiler.

Introduction

The interpretation of DNA profiles derived from mixtures with small amounts or degraded DNA is challenging due to difficulties in estimating mixture proportions and genotype combinations. Probabilistic genotyping software (PGS) based on the fully-continuous model is useful for interpreting these profiles by considering peak heights and implementing statistical models that explain stochastic effects such as peak height imbalance and allelic drop-out [1,2].

We previously developed the PGS "Kongoh" to analyze mixed DNA profiles typed by the AmpFlSTRTM IdentifilerTM Plus PCR Amplification Kit (Thermo Fisher Scientific, Waltham, MA) or the GlobalFilerTM PCR Amplification Kit (Thermo Fisher Scientific), which are used in Japan's police laboratories [3,4]. Kongoh performs a Monte Carlo simulation for deconvoluting the crime stain profile to assign the weights of all possible genotype combinations. However, the Monte Carlo simulation is not efficient in terms of computation and Kongoh cannot continuously estimate mixture proportions and degradation parameters for each contributor. In addition, Kongoh does not implement the allelic drop-in model.

To address these issues, we developed a new PGS "KongohPlus" based on a Markov chain Monte Carlo (MCMC) approach.

Materials and methods

KongohPlus was developed using R language (ver. 4.4.0). We used the Shiny package to build an interactive graphical user interface application. The MCMC models were written in the Stan codes and the CmdStanR interface was used to run the codes.

KongohPlus has 2 main functions: deconvolution of the crime stain profile and assignment of the likelihood ratio (LR). In the deconvolution, the MCMC is performed to assign the weights of all possible genotype combinations including the possibility of the allelic drop-out and drop-in. We used the Hamiltonian Monte Carlo algorithm as the MCMC method, which is adopted in the Stan. The parameters of the MCMC are mixture proportions per contributor, degradation

parameters per contributor, total peak height that reflects the total DNA amount, and locus-specific amplification efficiencies. These parameters are estimated continuously. The average allele-specific stutter ratios for back stutters, forward stutters, double-back stutters, and minus 2-nt stutters are fixed during the MCMC and determined based on the user-prepared experimental data. All possible genotype combinations are considered in each MCMC step based on the method proposed by Susik et al [5]; therefore, the number of warm-up steps and sampling steps are relatively small. The default numbers of warm-up steps and sampling steps are set to 1,200 and 300, respectively. The Rhat values (also known as the Gelman-Rubin diagnostics [6]) of the likelihood and each MCMC parameter are used for convergence diagnostics. After the deconvolution has been performed, the LR values can be assigned based on the deconvolution results and user-defined prosecution and defense hypotheses.

To determine the average allele-specific stutter ratios, 350 single-source profiles were prepared experimentally. DNA amounts were 1, 0.5, 0.25, 0.125, 0.063, 0.031, and 0.016 ng (n = 50 in each). PCR amplification was performed at 30 cycles using the GlobalFiler™ PCR Amplification Kit (Thermo Fisher Scientific). PCR products were then analyzed on an Applied Biosystems™ SeqStudio™ Genetic Analyzer (Thermo Fisher Scientific). The obtained data were analyzed using GeneMapper™ ID-X Software v1.6 (Thermo Fisher Scientific) with a detection threshold of 50 RFU.

We then analyzed 2-person mixtures (n = 24), 3-person mixtures (n = 24), and 4-person mixtures (n = 1) using KongohPlus as a pre-trial. Mixture ratios and DNA amounts of each mixture are shown in Table 1. For each mixed DNA profile, convergence diagnostics were performed based on each deconvolution result, and LR values were calculated for true contributors and 100 non-contributors which were randomly generated according to Japanese allele frequencies [7]. The computer specifications were Windows 10 64 bit, Intel Core i9-9900K CPU, 3.6 GHz, 32 GB RAM.

Results

The LR values of true contributors tended to increase as the DNA amount of the person of interest (POI) increased (Fig. 1). Except for one minor contributor of the 19:1 mixture with 0.125 ng DNA, the LR values of the true contributors were greater than 100,000, which can be considered as “strong support” for the proposition that the POI is a contributor [8].

The LR values of non-contributors tended to decrease as the DNA amount of the POI increased (Fig. 1). 65.4% of the LR values for non-contributors were less than 10^{-30}, and 58.7% of the LR values for non-contributors were zero because some observed peaks could not be explained under the proposition that the POI is a contributor. 0.97% of the LRs for non-contributors were greater than 1 (i.e., false positives), and the maximum LR was 239.

Convergence diagnostics were achieved for all experimental mixtures. All Rhat values for the likelihood and each MCMC parameter were less than 1.1, which typically indicates that the chains have likely converged. The computational time was approximate 20 sec in 2-person mixtures, approximate 10 min in 3-person mixtures, and approximate 11 hours in a 4-person mixture. The computational time of the 4-person mixture was reduced to 6 hours without considering double-back stutters.

Discussion

In this study, we developed a PGS "KongohPlus" to improve the previously developed PGS "Kongoh". We implemented the Hamiltonian Monte Carlo algorithm in KongohPlus to perform the MCMC, which is more efficient for estimating parameters for complex mixtures than the Monte Carlo simulation used in Kongoh. The implementation of MCMC resulted in a continuous estimation of mixture proportions and degradation parameters for each contributor, and convergence diagnostics were achieved for all experimental mixtures. The computational time was slightly reduced especially in the 4-person mixture comparing to Kongoh (30 sec in 2-person mixtures, 11 min in 3-person mixtures, and 17 hours in 4-person mixtures [4]). Therefore, KongohPlus is useful for interpreting DNA profiles derived from mixtures with small amounts or degraded DNA.

In the future, we will perform the developmental validation of KongohPlus preparing the experimental profiles typed by the GlobalFiler, which is generally used for actual caseworks in Japan. KongohPlus will be released on a GitHub web page.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 20K18981.

Conflict of interest statement

None.

References

[1] D. Taylor, J.-A. Bright, J. Buckleton, The interpretation of single source and mixed DNA profiles, Forensic Sci. Int. Genet. 7 (2013) 516–528.

[2] M.D. Coble, J.-A. Bright, Probabilistic genotyping software: An overview, Forensic Sci. Int. Genet. 38 (2019) 219–224.

[3] S. Manabe, C. Morimoto, Y. Hamano, S. Fujimoto, K. Tamaki, Development and validation of open-source software for DNA mixture interpretation based on a quantitative continuous model, PLoS One 12 (2017) 1–18.

[4] S. Manabe, T. Fukagawa, K. Fujii, N. Mizuno, K. Sekiguchi, A. Akane, K. Tamaki, Development and validation of Kongoh ver. 3.0.1: Open-source software for DNA mixture interpretation in the GlobalFiler system based on a quantitative continuous model, Leg. Med. 54 (2022) 101972.

[5] M. Susik, H. Schönborn, I.F. Sbalzarini, Hamiltonian Monte Carlo with strict convergence criteria reduces run-to-run variability in forensic DNA mixture deconvolution, Forensic Sci. Int. Genet. 60 (2022) 102744.

[6] L. Russell, S. Cooper, R. Wivell, Z. Kerr, D. Taylor, J. Buckleton, J.-A. Bright, A guide to results and diagnostics within a STRmix™ report, WIREs. Forensic Sci. 1 (2019) e1354.

[7] K. Fujii, H. Watahiki, Y. Mita, Y. Iwashima, T. Kitayama, H. Nakahara, N. Mizuno, K. Sekiguchi, Allele frequencies for 21 autosomal short tandem repeat loci obtained using GlobalFiler in a sample of 1501 individuals from the Japanese population, Leg. Med. 17 (2015) 306–308.

[8] Scientific Working Group on DNA Analysis Methods, Recommendations of the SWGDAM ad hoc working group on genotyping results reported as likelihood ratios, (2018).

Table

Table 1. Experimental DNA profiles

Number of contributors	Mixture ratio	DNA (ng)
2	1:1	1, 0.5, 0.25, 0.125
	3:1	
	9:1	
	19:1	
3	1:1:1	1, 0.5, 0.25, 0.125
	3:2:1	
	9:9:2	
	8:1:1	
4	4:3:2:1	0.25

Figure caption

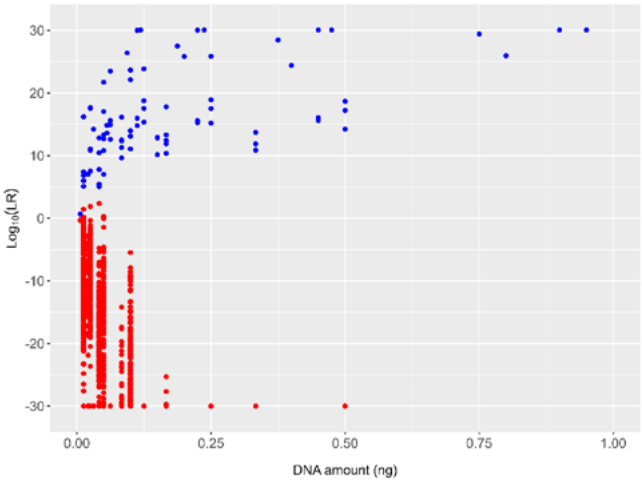


Figure 1. Relationship between likelihood ratios (LRs) and DNA amounts of the person of interest. Blue circles represent $\log_{10}$(LR) values for true contributors, and red circles represent $\log_{10}$(LR) values for non-contributors. LR values < 10^{-30} are plotted at $\log_{10}$(LR) = -30. The DNA amount for non-contributors was taken as the minimum value among the true contributors in each profile.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 155-159
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Bridging expertise and data-driven approaches in forensic analysis: Validation of the DNA Decision Support Tool

Corina C.G. Benschop[1,*], Maartje Luijsterburg[1], Brechtje Hoegen[1], Claire Wagenaar[1], Charissa van Kooten[1], Margreet van den Berge[1], Francisca E. Duijs[1], Pieterjan Sjoukema[1], Rolf Ypma[1]

[1]*Netherlands Forensic Institute, The Hague, The Netherlands*

** Corresponding author at: c.benschop@nfi.nl*

Abstract
This study highlights the validation of the Search dashboard within the DNA Decision Support Tool for casework application. The tool provides reporting officers (ROs) with valuable insights into actions taken and outcomes from previous DNA casework. This tool helps ROs to make informed decisions in current cases and enables to better substantiate their choices in the DNA profile interpretation process.

Keywords
Forensic science, DNA profile interpretation, Rework, Casework data, DNA Decision Support Tool

Introduction to the DNA Decision Support Tool

Reporting officers (ROs) make key decisions throughout the process, from the initial DNA profile to the final report. For example, whether to perform actions such as to submit or search within the criminal database, or to conduct a weight of evidence calculation. Additionally, ROs may choose to apply rework to enhance the DNA profiling result, improving the chances of successful outcomes in one of these actions. Examples of rework include generating additional PCR replicates, purifying the DNA extract, adjusting PCR input (increasing or diluting), re-analysing the DNA profile using lower analytical thresholds, or employing a different STR typing

kit. The decisions are influenced by a combination of guidelines, available resources, expert knowledge and experience, and can impact the time spent on a case, the resources used, and the final reported outcome.

To our knowledge, no laboratory systematically utilises historical casework data to assist in these decisions. To address this gap, we have developed a DNA Decision Support Tool [1]. An overview of the components and data in the DNA decision support tool is shown in Fig. 1. The user interface contains various dashboards and this study focusses on the validation of the Search dashboard. In its current version (July 2024), the tool allows users to upload an initial casework DNA profile, automatically assign a profile category (e.g. "singular profile"), search for profiles with similar characteristics (e.g. peak height, allele count, degradation slope), and review the actions taken on these similar profiles and their outcomes in previous cases. Additionally, it estimates the number of additional peaks that could be obtained if the initial profile would be analysed using lower analytical thresholds, which can help in deciding whether or not to use this type of rework.

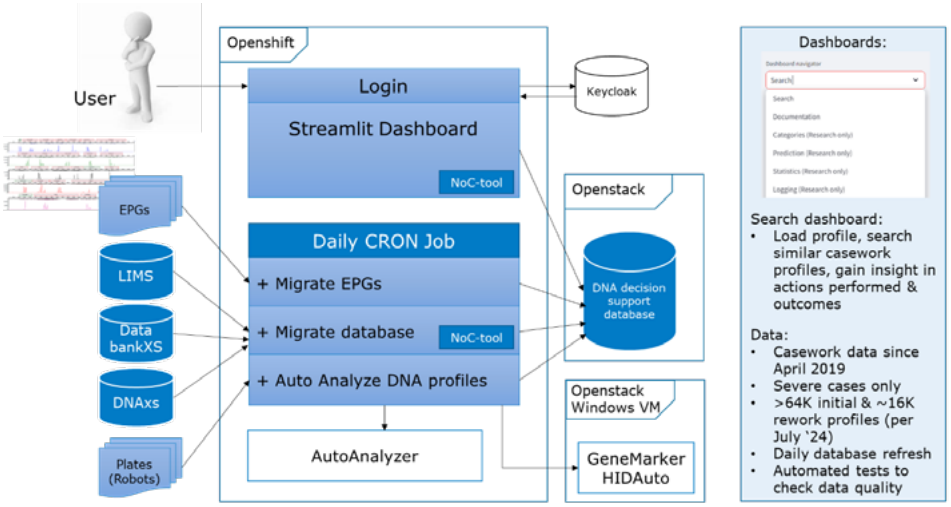


Fig. 1. Overview of the components of the DNA Decision support tool.

Validation results

Validation of the Search dashboard demonstrated that similar profiles are obtained, and actions, rework and outcomes are shown. The study approved amongst others reproducibility, repeatability, precision (correct information is shown and – when searching for similar profiles - the uploaded profile is ranked first with 100% similarity), monster tracking (database is up to date with freshly generated data from the day before), applicability of the software (tool applies to PowerPlex Fusion 6C profiles generated within our institute), accessibility of the software (only accessible to persons who are allowed to access casework data), inclusion of automated software tests (quality control of the database), robustness test (simultaneous use of the tool by multiple users is possible). Two features were examined in further detail:

1) DNA profile categories

Based on the expected action with a profile, ten profile categories were defined (Table 1A). Categories were automatically assigned to DNA profiles in the database based on their profile characteristics (Table 1B). Table 1C shows the distribution of previous casework data per category and rework performed. Manually assigned profile categories were compared to the automatically assigned categories (Table 1D).

Results show that most profiles fall into their expected category (>85%, except category 7, Table 1D) and follow the expected trends for e.g. rework and database submission. For instance, categories 2, 6 and 8 with a low template component were, expectedly, most often subjected to rework.

Table 1. Overview of the ten profile categories. A) Expected actions per category. B) Characteristics that define categories. C) Number and percentage of casework profiles per category and for which rework was performed. D) Percentage of profiles for which manual assessment and automated classification yielded the same profile category.

	DNA profile category and description [a]									
A	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Majority of DNA profiles expectedly used for	**Insufficient informative**	**Low template (LT)/ degraded 1p**	**1p**	**High template (HT) 2p**	**1 major & minor**	**1 major & minor(s) subthreshold**	**3p-4p HT, No clear major**	**2p-4p LT, No clear major**	**Multiple majors (<4) & mINor(s)**	**Many donors (>4), >3 majors**
Rework	No	Yes	No	No	No (major) No/Yes (minor(s))	No (major) Yes (minor(s))	No	Yes	No	No
Inference major	No	No	Only if some minor peaks	No	Yes	Yes	No	No	No	No
DNA database (DB) submission	No	No	Yes	Yes	Yes	Yes (major) No (minor(s))	No	No	No	No
LR based DB search	No	Yes/No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Weight of evidence (WoE) calculation	No	Yes/No	Yes	Yes	Yes	Yes (major) No	Yes	Yes	Yes	No
TopDown approach for WoE	No	No	No	No	No	Possibly	No	No	Yes	No
B **Characteristics that define categories for PPF6C DNA profiles**										
PCR input (uL)	7.5 (max)									
Maximum allele count (MAC)			2-4 [b]	3-5	≥3		5-9	4-9	≥6 [b]	≥10 [b]
Total allele count (TAC)	≤5	≤46							≥65 [b]	≥110 [b]
TAC+TAC subthreshold	≤10 or ≥80	≤50 or ≥100								
TAC subthreshold				≤5	≤5	≥5	≤9	≥5		
Number of contributors (estimated)		1-2	1 [b]	2	2-3	≥2	3-4	2-4	≥3 [b]	≥5 [b]
All Peaks >stochastic threshold (ST)	FALSE									
Avg. Peak height >ST	FALSE	FALSE [b]		TRUE	TRUE		TRUE	FALSE		
#Type I loci (based on LoCIM [d])		0-10					≤6	≤5	≤6	
#Type III loci (based on LoCIM)			0-8	≥10	≤8	≤12	≥8		≥8	
#inferred alleles			20-47		20-47	≥20				
#remaining peaks after inference			0-15		≥10	≥3				
Degradation slope				≥0.6					≥0.7	
#locus drop-outs (autosomal)	18-23	≤20 [b]	0-5	0	0	0-2	0-2		0-2	
C **Initial profiles per category** [c]	1637	14174	23721	503	2724	8878	826	7856	946	365
	3%	22%	37%	1%	4%	14%	1%	12%	1%	1%
Initial profiles (at stain identification level) with rework [c]	128	2133	1046	6	120	790	26	1607	36	4
	2%	33%	16%	0%	2%	12%	0%	25%	1%	0%
D **Assigned category as expected (subset manually examined)**	100%	95%	100%	90%	86%	100%	16%	95%	90%	90%

[a] Profile order as programmed: 1, 3, 2, 4, 5, 6, 8, 7, 9, 10. Profiles that meet the criteria of both category 2 and 3 will be assigned to category 3.
[b] Category 2, 3, 9 and 10: and/or criteria for these two/three parameters. [c] Not all profiles are assigned to a category (5% uncategorized).
[d] LoCIM information in Benschop et al. Forensic Sci. Int. Genet. 11 (2014) 154-165.

2) Number of subthreshold peaks

Raw DNA profiles are automatically reanalysed using lower analytical thresholds. Automated and manual analysis were compared and show similar trends for the subthreshold maximum and total allele count (data not shown). Discrepancies are explained based on the data (software version, file availability, etc.).

Conclusions & future work

The Search dashboard of the DNA Decision Support Tool was found valid as a useful resource for gaining insight into current DNA profiles and previous casework. The profile categories assigned to each DNA profile can be used as a predictor for

the appropriate action to take. Additionally, the automatically displayed number of expected subthreshold peaks provides useful information for determining whether re-analysis with lower analytical thresholds would be beneficial.

This tool allows for a more data-driven approach to DNA profile interpretation, ensuring efficient and optimal use of available resources.

Future development of the DNA Decision Support Tool will focus on predicting success rates of rework for actions such as DNA database submission. The tool, along with the underlying database, also facilitates research inquiries aimed at monitoring, optimising or automating laboratory processes.

Acknowledgments

We are grateful for the support of the Innovatieplatform SRK and we thank the colleagues who contributed to the development of the DNA Decision Support Tool.

Conflict of interest statement

Authors have no conflict of interest to declare.

References

[1] C.C.G. Benschop, C. van Kooten, D. Zandstra, P. Sjoukema, F.E. Duijs, M. van den Berge, C. Schepers, J. Mandersloot, R. Ypma. Using previous DNA casework data to aid decision making in the process of DNA profile interpretation. Forensic Science International: Genetics Supplement Series 8 (2022) 251-253.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 161-168
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Exploring STRs as an approach to continental biogeographical ancestry inference

Lucía Casanova-Adán[1*], Javier González-Bao[1], Ana Mosquera-Miguel[1], Adrián Ambroa-Conde[1], Jorge Ruiz-Ramírez[1], Amaia Cabrejas-Olalla[1], Miguel Boullón-Cassau[1], Ana Freire-Aradas[1], Amelia Rodríguez-López[1], Christopher Phillips[1,2], María Victoria Lareu[1], María de la Puente[1*]

[1]*Forensic Genetics Unit, Institute of Forensic Sciences, Universidade de Santiago de Compostela, Spain*
[2]*King's Forensics, Faculty of Life Sciences and Medicine, King's College, London, UK*

** lucia.casanova.adan@usc.es / m.delapuente.vila@gmail.com*

Abstract

Bio-geographical ancestry (BGA) inference provides valuable information to guide investigators when there is no database hit, or no suspects are available for comparison. This approach is usually carried out through the analysis of panels that collect different ancestry informative markers (AIMs), centering on SNPs, but also including microhaplotypes or InDels, as markers showing different allele frequencies between populations.

This study aims to explore BGA informativeness for the different STR kits available for human identification (HID), including CE and MPS panels. We explore the degree of differentiation at continental level achieved by the panels, using the HGDP-CEPH panel as sample set, and performed a comparative evaluation of the results to understand their *limitations as an initial approximation to BGA, in situations where it is necessary to avoid sample depletion, or all DNA has been used.*

Keywords

Bio-geographical ancestry, STRs, structure, HID panel

Introduction

The use of STRs for biogeographical ancestry (BGA) analysis is limited due to their higher mutation rate and degree of polymorphism, despite being the standard

markers in the forensic genetics field. Nowadays, the arrival of massively parallel sequencing (MPS) and increasingly sophisticated capillary electrophoresis (CE) systems enable larger STR multiplexing capabilities, providing greater discriminatory power and informativeness.

We consider the use of the above kits for a dual purpose: human individual identification (HID) and an initial approximation to a continental BGA, which can help avoid sample depletion from multiple analyses, or when all DNA has been used and only STR profiles are available. Moreover, STRs - as multiallelic markers - are easier to deconvolute in comparison to AIM-SNPs, prompting the possibility of BGA inference of the major and minor components in simple DNA mixtures. *Therefore, this study aims to explore BGA informativeness for different STR commercial kits available for human identification, including CE and MPS panels.*

Material studied, methods, techniques

Dataset generation

Genotypes from 578 HGDP-CEPH panel individuals, corresponding to five major biogeographical regions and aiming to exemplify the global variation of the human populations at a continental level, including 103 SubSaharan Africans (AFR), 158 Europeans (EUR), 226 East Asians (EAS), 27 Oceanians (OCE) and 64 Native Americans (NAM); were gathered from previous works [1–3]one of the most widely used forensic STR multiplexes, plus five recently introduced European Standard Set (ESS in the case of STRs and whole genome sequencing public data [4]and relationships between, different populations. We present 929 high-coverage genome sequences from 54 diverse human populations, 26 of which are physically phased using linked-read sequencing. Analyses of these genomes reveal an excess of previously undocumented common genetic variation private to southern Africa, central Africa, Oceania, and the Americas, but an absence of such variants fixed between major geographical regions. We also find deep and gradual population separations within Africa, contrasting population size histories between hunter-gatherer and agriculturalist groups in the past 10,000 years, and a contrast between single Neanderthal but multiple Denisovan source populations contributing to present-day human populations.","container-title":"Science","DOI":"10.1126/science.aay5012","issue":"6484","note":"publisher: American Association for the Advancement of Science","page":"eaay5012","source":"science.org (Atypon in the case of SNPs.

A total of six different panels were considered (Table 1), containing the representative set of markers from different MPS and CE HID commercial kits.

Table1: Generated panels, including their designation code, the commercial name of the kit, the type of sequencing technology, and the total number of autosomal markers.

Panel	Kit commercial name	Technology	Number of markers
A	ForenSeq SignaturePrep	MPS	121 (27 STRs+94 SNPs)
B	Precision ID GlobalFiler NGS STR Panel V2	MPS	30 STRs
C	IDseek OmniSTR	MPS	28 STRs
D	ForenSeq MainstAY	MPS	27 STRs
E	PowerSeq 46GY System	MPS	22 STRs
F	GlobalFiler PCR Amplification kit	CE	21 STRs

Population structure analysis

The different panels were uniformly processed using the systematic Bayesian clustering approach STRUCTURE v.2.3.3 [5]each of which is characterized by a set of allele frequencies at each locus. Individuals in the sample are assigned (probabilistically. For each panel, a total of 5 iterations at K=5 were performed with 100,000 burnin steps and 100,000 MCM steps, with correlated allele frequencies under the Admixture model and the no POPFLAG option. Ancestry membership proportion plots were constructed using CLUMPAK through the StructureSelector website [6,7].

Data processing was carried out in the RStudio (v.4.3.2) environment, with "dplyr" and "tidyr" packages, and common spreadsheets. During this stage, individuals with missing values at any marker were excluded to avoid bias and ensure consistency of the results.

Results and discussion

Figure 1 represents ancestry membership proportion plots generated using the STRUCTURE + StructureSelector pipeline for each panel, sorted in decreasing number of autosomal markers. The analysis of HID kits with a larger number of markers has a beneficial effect in distinguishing between continental clusters, highlighting the effect of the 94 HID-SNPs in panel A. A greater level of noise in

the clusters is observed when the number of markers drops below thirty, and this decline becomes more appreciable as the number of markers decreases.

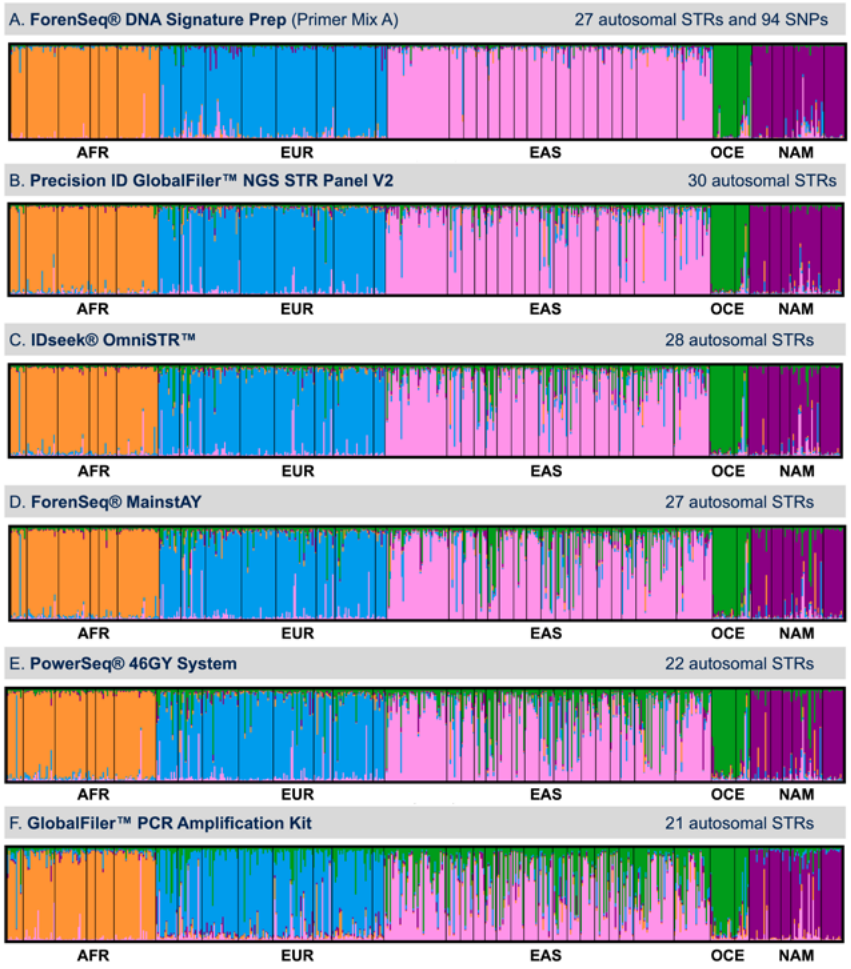


Figure 1: STRUCTURE + StructureSelector analyses obtained at K=5 for each of the panels used in the study, shown in decreasing number of markers. Clusters are colored according to Rosenberg et al. 2002 [8].

From the co-ancestry values obtained, two different criteria for ancestry assignment were considered: (i) the individuals belong to a certain population if their

membership coefficient exceeds or equals the 0.8 threshold (>0.8), and (ii) the individuals belong to the population showing the highest membership coefficient (major). Results from both classifications are shown on Table 2, listing percentages of correctly classified individuals under both conditions.

Table 2: Percentage of correct classification rates for each panel based on two criteria: (>0.8) percentage of individuals with an accurate group membership coefficient greater than 0.8 and (major) percentage of individuals where the major membership component corresponds to their actual population group. $\bar{x}$: mean percentage of correct classification values across populations.

Panel	Criteria	AFR	EUR	EAS	OCE	NAM	$\bar{x}$
A	>0.8	97.83%	88.81%	93.09%	92.31%	83.72%	91.15%
	major	100.00%	99.30%	99.47%	100.00%	97.67%	99.29%
B	>0.8	96.15%	87.97%	79.23%	100.00%	96.67%	92.01%
	major	100.00%	98.50%	94.54%	100.00%	100.00%	98.61%
C	>0.8	93.10%	80.31%	67.20%	100.00%	97.14%	87.55%
	major	100.00%	92.13%	89.25%	100.00%	97.14%	95.70%
D	>0.8	92.05%	81.89%	65.59%	100.00%	97.14%	87.33%
	major	98.86%	96.85%	87.63%	100.00%	97.14%	96.10%
E	>0.8	92.47%	68.28%	49.00%	100.00%	89.47%	79.84%
	major	98.92%	89.66%	79.00%	100.00%	97.37%	92.99%
F	>0.8	83.16%	62.76%	33.65%	81.82%	87.80%	69.84%
	major	98.95%	85.52%	62.50%	100.00%	97.56%	88.91%

The ForenSeq Signature Prep kit (A) features the highest number of markers, a total of 121 (27 autosomal STRs and 94 SNPs) and yields the most optimal results. When applying the 0.8 threshold, the best outcomes are seen in African, East Asian, and Oceanian populations, with the lowest performance observed in Native Americans. In terms of the highest membership coefficient, both Africa and Oceania stand out with a 100% of accurate classifications, a finding that was also replicated in other analysed kits.

The Precision ID GlobalFiler NGS STR Panel V2 (B) ranks second in number of markers, with a total of 30 autosomal STRs. For both criteria, the best results are found in African, Oceanian and Native American populations.

In third place, with 28 autosomal STRs, the NimaGen OmniSTR kit (C) follows the previously established trend of better performance for African, Oceanian, and Native American populations under both criteria. However, it shows weaker results for Europe and East Asia, where assignment consistency begins to decline.

ForenSeq MainStAY kit (D) comprises 27 autosomal STRs. Similar to the NimaGen kit, it demonstrates strong classification performance for African, Oceanian, and Native American population data, with slightly better results for the European population within the 0.8 threshold, despite sharing all markers but SE33 with the previous panel.

The PowerSeq panel (E) contains 22 autosomal STRs, and its results follow the previously established classification trend, with high prediction accuracy for African, Oceanian, and Native American populations. In contrast, European and East Asian populations yield less conclusive results, particularly East Asian, where erroneous predictions exceed the 50% with the 0.8 threshold.

GlobalFiler CE kit (F), which includes 21 autosomal STRs, exhibits the poorest HID kit performance overall. While it still provides a relatively good classification for African, Oceanian, and Native American populations, the results for Europeans and East Asians are weaker, especially for East Asian, as previously noted.

Concluding remarks

The number of markers likely plays an additive role, with more markers leading to higher rates of correctly classified individuals. This underscores the advantage of MPS kits, which offer a higher multiplex capacity, in contrast to CE, whose capacity may not be robust enough for the demands of BGA inference.

African, Oceanian, and Native American populations achieved more reliable classification results compared to Europeans or East Asians, especially depending on the number of available markers.

Nevertheless, the results shown by the largest panels indicate a certain level of information that can be used for inferring BGA at a continental level for investigative purposes, providing that no extra sample is available and understanding its limitations.

Acknowledgments

LCA is supported by a predoctoral grant funded by "Fundación Caixa Rural Galega Tomás Notario Vacas" in collaboration agreement with the Universidade de Santiago de Compostela. JGB is supported by predoctoral grant "Ayudas para la Formación del Profesorado Universitario" (reference FPU22/02683) funded by Ministerio de Universidades from Gobierno de España. MBC is supported by the "Axudas de apoio á etapa predoutoral" funded by the Consellería de Cultura, Educación, Formación Profesional e Universidades from Xunta de Galicia, Spain (ED481A-2024-163). MVL is supported by grant PID2019–107876RB-I00 funded by the Ministerio de Educación, Cultura y Ciencia, Spain (MCIN/AEI/ 10.13039/501100011033) and grant PID2022-141224OB-I00 funded by MCIN/AEI/ 10.13039/501100011033 and "ERDF A way of making Europe". MdP is supported by grant IJC2020–042638-I funded by the Gobierno de España MCIN/AEI/10.13039/501100011033 and the European Union "NextGenerationEU/PRTR".

Conflict of interest statement

The authors have declared no conflict of interest

References

[1] Phillips C, Fernandez-Formoso L, Garcia-Magariños M, Porras L, Tvedebrink T, Amigo J, et al. Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel. Forensic Sci Int Genet. junio de 2011;5(3):155-69.

[2] Phillips C, Kind S, Fernandez-Formoso L, Gelabert-Besada M, Carracedo A, Lareu MV. Global population variability in Promega PowerPlex CS7, D6S1043, and Penta B STRs. Int J Legal Med. septiembre de 2013;127(5):901-6.

[3] Phillips C, Fernandez-Formoso L, Gelabert-Besada M, García-Magariños M, Amigo J, Carracedo A, et al. Global population variability in Qiagen Investigator HDplex STRs. Forensic Sci Int Genet. enero de 2014;8(1):36-43.

[4] Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, et al. Insights into human genetic variation and population history from 929 diverse genomes. Science. 20 de marzo de 2020;367(6484):eaay5012.

[5] Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. junio de 2000;155(2):945-59.

[6] Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour. septiembre de 2015;15(5):1179-91.

[7] Li YL, Liu JX. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. Mol Ecol Resour. enero de 2018;18(1):176-7.

[8] Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic Structure of Human Populations. Science. 20 de diciembre de 2002;298(5602):2381-5.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 169-173
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Development of a new microhaplotype panel for ancestry inference

Pedro Rodrigues[1,2,3,4,*], Nádia Pinto[2,3,5], Maria João Prata[2,3,4], Claus Børsting[1], Leonor Gusmão[6], Vania Pereira[1]

[1]*Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark*

[2]*Instituto de Investigação e Inovação em Saúde (i3S), Porto, Portugal*

[3]*Institute of Molecular Pathology and Immunology, University of Porto (IPATIMUP), Porto, Portugal*

[4]*Faculty of Sciences of the University of Porto (FCUP), Porto, Portugal*

[5]*Centre of Mathematics of the University of Porto, Porto, Portugal*

[6]*DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil*

** Corresponding author at: Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, pedro.rodrigues@sund.ku.dk*

Abstract

Over the last decade, microhaplotypes (MHs) have proven to be a reliable complement or even an alternative to other types of genetic markers for several forensic applications, including biogeographical ancestry inference. In this study, we developed a new ancestry MH panel, named MHappaMundi, to separate 5 metapopulations (Sub-Saharan Africans, Europeans, South Asians, East Asians, and Native Americans). The MHs selection involved locus-by-locus AMOVA between each pair of population groups, using profiles from the 1000 Genomes database. The final selection included 100 ancestry MH loci comprising 298 SNPs. Panel validation involved sensitivity tests with DNA inputs of 0.2, 0.5, 1, and 5 ng, PCR and library preparation with full and half reagent volumes, and PCR with 21 and 24 cycles. All tests were performed in duplicates. Six MH loci failed to amplify or exhibited many sequencing errors, and were excluded from analysis. As for the remaining loci, using as little as 0.2 ng of DNA input and 21 PCR cycles was enough to consistently produce ≥100 reads per locus.

Keywords

Microhaplotypes, MPS, NGS, biogeographical ancestry inference, forensic genetics, population genetics

Introduction

Microhaplotypes (MHs) are a novel class of genetic markers, that consist in haplotypes defined by two or more polymorphisms (SNPs and/or indels) within a fragment of less than 300 bps [1]. Several publications report the performance of selected and newly discovered MHs, as well as the development of panels for different forensic purposes (see [2] and references therein for examples). However, MH panels for biogeographic ancestry inference are scarce and the existing ones were selected with the purpose of separating subpopulations within a continent [3], or lacked laboratory testing [4,5].

Our study details the development of a new ancestry informative MH panel named MHappaMundi. The MHs were chosen for their effectiveness in distinguishing five population groups: Sub-Saharan Africans (AFR), Europeans (EUR), South Asians (SAS), East Asians (EAS), and Native Americans (NAM). This work involved a validation study where the performance of the panel was tested under different conditions.

Materials and methods

Selection of MH loci for the ancestry panel

The ancestry informative MHs were selected from the MicroHapDB [6], a database containing information for 3053 MHs (assessed on 30 September 2024). Genotype data from 2958 profiles from 26 populations were retrieved from the 1000 Genomes database [7].

The populations were grouped into 5 main groups: AFR, EUR, SAS, EAS, and NAM. Pairwise FST values were calculated for each MH through locus-by-locus AMOVA in Arlequin ver 3.5.2 software [8]. A ranking of MH loci was created for each pair of metapopulations based on the FST values obtained. MHs with the highest FST were included in the final panel, ensuring similar overall FST distances between all population pairs. A total of 100 ancestry informative MHs comprising 298 SNPs were selected. In silico panel performance was investigated by STRUCTURE v.2.3.4 [9] analysis with the 1000 Genomes dataset. STRUCTURE analysis was performed using a burn-in of 100,000 steps and 10,000 MCMC steps applying the admixture model with correlated allele frequencies.

Methodology

Primers were designed with the online available tool Ion AmpliSeq™ Designer (https://ampliseq.com/; ThermoFisher Scientific, TFS, Waltham, MA, USA) on 'Made-to-order' section for fragments up to 375 bps.

Different DNA inputs (0.2, 0.5, 1, and 5 ng), PCR and library preparation reagent volumes [full (FV) and half volume (HV)], and number of PCR cycles (21 and 24) were tested in duplicates using the control DNA 007.

Libraries were manually prepared following the manufacturer's protocol, and adjusting DNA input, reagent volume, and the number of PCR cycles when necessary. Samples were loaded onto an Ion 530™ sequencing chip (2 chips containing 24 samples each in the validation study, and 2 chips containing 30 samples each in the population study) and sequenced on the Ion GeneStudio S5™ system (Thermo Fisher Scientific) using the Precision ID S5™ Sequencing Kit (Thermo Fisher Scientific) and Ion 530™ Chip Kits (Thermo Fisher Scientific).

MHinNGS [10] was used to perform noise filtering and genotype calling.

Results

To evaluate the performance of the selected MHs, we performed a STRUCTURE analysis using genotype data from populations from the 1000 Genomes Project. Populations were grouped in 6 groups: African Americans, Sub-Saharan Africans, Europeans, South Asians, East Asians, and Native Americans. This latter group was subdivided in two: the first subgroup included individuals from the admixed American populations of Peru and Mexico who exhibited a Native American component exceeding 70% [11], and the second subgroup included the remaining individuals from both countries, as well as from Puerto Rico and Colombia. The STRUCTURE results (Figure 1), assuming K = 5, show a clear discrimination of the populations according to geographical regions, suggesting a large potential of the marker set to capture ancestry components from the corresponding 5 metapopulations.

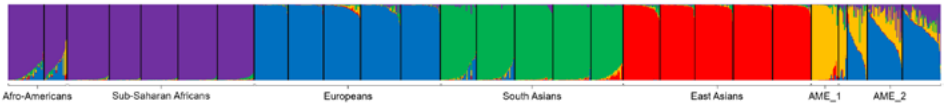


Figure 1. STRUCTURE plot, assuming K=5, with all the available profiles in 1000 Genomes Project. Individuals from Peru and Mexico, who exhibited more than 70% Native American ancestry [11] were categorized distinctly as 'AME_1', and all the remaining American individuals were categorized as 'AME_2'.

Sensitivity test results (Figure 2) showed few differences in the average read counts per locus between FV and HV, unless for DNA inputs lower than 1 ng. Overall, higher average read counts per locus were obtained with 24 PCR cycles (around 25% more reads), although locus imbalances in these samples were higher (standard deviations were 3.6, 3.5, 2.1, and 1.2 times higher for 0.2, 0.5, 1, and 5 ng of DNA input, respectively). Based on the sensitivity tests, 0.2 ng DNA and 21 PCR cycles generated MH profiles with at least 100 reads per locus. Six of the 100 MH loci were excluded from further analyses due to failed amplification or sequencing errors.

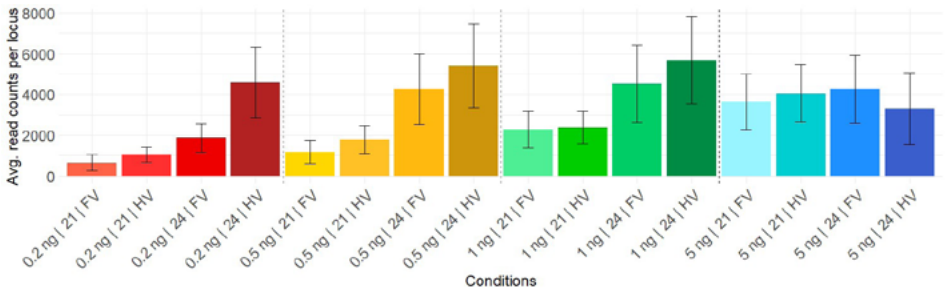


Figure 2. Average number of reads per locus using the tested conditions (DNA input | PCR cycles | Reagent volumes). Standard deviations are shown. Duplicates were merged and averaged.

Conclusion

- In silico tests demonstrated good performance of the MHappaMundi panel in 1000 Genomes samples;
- Sensitivity tests showed that using as little as 0.2 ng of DNA input and 21 PCR cycles consistently produced ≥100 read counts per locus;
- Further sequencing of samples from different ethnic backgrounds will shed light on the ancestry inference ability of the MHappaMundi.

Acknowledgments

PR is funded by a FCT—Fundação para a Ciência e a Tecnologia—doctoral grant 2022.11825.BD.

Conflict of interest statement

The authors declare no conflict of interest.

References

[1] Kidd, K. K. et al. (2013). Microhaplotype loci are a powerful new type of forensic marker. Forensic Sci. Int. Genet. Suppl. Ser. 4(1): e123-e124.

[2] Oldoni, F. et al. (2019). Microhaplotypes in forensic genetics. Forensic Sci. Int. Genet. 38: 54–69.

[3] Zou, X. et al. (2022). Screening and selection of 21 novel microhaplotype markers for ancestry inference in ten Chinese subpopulations. Forensic Sci.Int. Genet. 58, 102687.

[4] Bulbul, O. et al. (2018). Ancestry inference of 96 population samples using microhaplotypes. Int. J. Legal Med. 132, 703-711.

[5] Huang, S. et al. (2022). Inferring bio-geographical ancestry with 35 microhaplotypes. Forensic Sci. Int. Genet. 341, 111509.

[6] Standage, DS. et al. (2020). MicroHapDB: A Portable and Extensible Database of All Published Microhaplotype Marker and Frequency Data. Front. Genet. 11: 781.

[7] The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature 526, 68-74.

[8] Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10(3): 564-567.

[9] Pritchard, J. K. (2000). Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

[10] Jønck, C.G. & Børsting, C. (2022). Introduction of the Python Script MHinNGS for Analysis of Microhaplotypes. Forensic Sci. Int. Genet. Suppl. Ser. 8: 79–81

[11] Pereira, V. et al. (2020). Evaluation of the Precision of Ancestry Inferences in South American Admixed Populations. Front. Genet., 11, 966.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 175-182
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Integrating physical anthropological techniques and emerging methodologies using a quasi-landmark mesh to decode the upper facial region from DNA

Belén Navarro López[1,2,3*], Victoria Suárez-Ulloa[4], Franziska Wilke[5], Miriam Baeta[1,2,3], Rubén Martos-Fernández[6], Susan Walsh[5], Begoña Martínez-Jarreta[7], Susana Jiménez[8], Marian M. de Pancorbo[1,7]

[1]BIOMICs Research Group, Lascaray Research Center, University of the Basque Country (UPV/EHU), Vitoria-Gasteiz, Spain; [2]Department of Zoology and Animal Cellular Biology, Faculty of Pharmacy, UPV/EHU, Vitoria-Gasteiz, Spain; [3]Bioaraba Health Research Institute, Vitoria-Gasteiz, Spain; [4]Panacea Cooperative Research S. Coop, Castilla y León, Spain; [5]Indiana University Indianapolis, Indianapolis, United States; [6]Department of Legal Medicine, Toxicology, and Physical Anthropology University of Granada, Spain; [7]Faculty of Medicine, University of Zaragoza, Zaragoza, Spain; [8]Department of Pathology and Surgery, University of Miguel Hernández, Alicante, Spain

**Corresponding author at: belennavarrolopez1997@gmail.com*

Abstract

Genetic identification currently relies on comparing unknown DNA samples against reference profiles in forensic databases or indubitable individuals. However, when no match is available, these cases may remain unsolved. In such circumstances, advances in Forensic DNA Phenotyping (FDP), that seeks to determine externally visible characteristics (EVCs) from DNA, could assist in identification efforts. Among EVCs, facial shape is of particular interest. While genome-wide association studies (GWAS) have identified potential SNPs linked to facial traits, these vary between populations and methodology. Addressing this issue, the present study has selected a set of candidate SNPs and assessed their correlation with facial phenotype in a Spanish population, focusing on the upper facial region, which includes the highly informative ocular area. From a collection of 412 individuals, two strategies were performed: one based on traditional anthropometric measurements and indexes with Pearson/Spearman and Chi-squared analyses, and another using a mesh of quasi-landmarks with canonical correlation analyses. Results revealed significant associations between several SNPs and upper facial metrics, although the SNPs and regions identified differed between the two methods and in part, to previous published results. These findings underline the importance of methodology and validation in population groups in

facial description when conducting genetic studies, and most notably, when considering forensic applications.

Keywords

Forensic DNA Phenotyping, facial morphology, correlation study, candidate SNPs.

Introduction

Genetic identification is a key tool in forensic biology, allowing the identification of individuals through genetic profiles comparison. However, unresolved cases may arise when no match is found in genetic databases or among persons involved in the investigation. Recent advances have led to the development of FDP, which aims to predict EVCs from DNA, providing valuable information about an individual's physical appearance for identification purposes [1].

The prediction of facial shape is particularly important due to its recognizability, making it a highly desirable goal for FDP. Despite the complexity of identifying genes involved in facial morphology, recent GWAS have identified potential SNPs associated with facial features [2–18]. Nevertheless, findings from these studies often show inconsistencies, largely due to differences in population samples and analytical strategies. However, advancements in 3D imaging systems have facilitated comprehensive evaluation of global and local variation, enabling different approaches to analyze facial morphology and perform association studies between genetic markers and facial traits [4, 19, 20].

Given the observed variability in results among previous investigations, our research focuses on analyzing candidate SNPs within a Spanish population to evaluate their correlation with facial traits, particularly in the upper facial region, including the highly informative ocular area. We employed two approaches to assess the correlation: a traditional physical anthropology method and an innovative quasi-landmark mesh technique. By comparing these methods, our research seeks to identify the optimal approach for comprehensive analysis, and validate previously published SNPs in a Spanish population.

Material studied, methods, techniques

A total of 412 individuals from a Spanish population were studied, with ethical approval (M10_2021_143) from the Ethics Committee for Research on Human Subjects of the University of the Basque Country. All participants provided written informed consent and completed a form detailing variables potentially influencing facial shape (sex, age, weight, height, surgery, and pathologies/traumas).

Additionally, saliva samples were collected in triplicate using sterile swabs, and 3D facial images were obtained using a 3D scanner (Academia 3D/20, Creaform) following established protocols [21]. Post-acquisition, 3D images were processed using the ACADEMIA Software Bundle.

DNA extraction was performed using the DNA Purification System Puregen™ (Gentra System, Inc.). A set of 116 candidate SNPs previously related to facial morphology [2–18] was analyzed using Fluidigm (Fluidigm Corp.) and SNaPshot minisequencing (Applied Biosystems) technologies. SNPs were coded according to the additive genetic model (AA = 0, Aa = 1, aa = 2).

Physical anthropology approach

3D images were imported into Skeleton-ID software [22] to place five cephalometric landmarks (en', ex', ft', fz', and zy'), used to calculate six linear measurements and three proportionality indices (face width, superior facial width, minimum frontal width, intercanthal width, biocular width, eye fissure width, forehead-face width index, intercanthal index, and palpebral fissure length index). These metrics were analyzed as both continuous and categorical variables. Associations with continuous variables were evaluated using Pearson or Spearman correlation tests, depending on data normality. For categorical variables, Chi-square tests ($\chi 2$) were employed. Variables were classified into three categories (smaller, normal, increased) based on their mean ± 0.5x standard deviation. All analyses were conducted in RStudio [23].

Quasi-landmark mesh approach

A quasi-landmark mesh was aligned to all 3d images using the MeshMonk pipeline [19, 24]. Facial scans were symmetrized by averaging each image with its reflection, and aligned in space using Procrustes superimposition. Covariate effects (sex, age, height, and weight) were removed using partial least-squares regression. Faces were then divided into 63 segments following the labels of Sero et al. [25], with 15 regions corresponding to the upper face. Each segment underwent generalized Procrustes Superimposition, followed by principal component (PCA) and Horn's parallel analysis to capture key variance and reduce dimensionality. Associations between SNPs and phenotypic variation in each of the 15 facial segments, represented as PCs, were evaluated through canonical correlation analysis (CCA). Additionally, a false discovery rate according to Benjamin-Hochberg (FDR) threshold of 0.05 was applied. Normal displacement maps were plotted for significant SNPs. All code was adapted from prior publications using Matlab 2023b [16, 26].

Results

Our results identified statistically significant associations for eight out of 116 SNPs with the metrics analyzed using traditional physical anthropology methods (Table 1). These SNPs were selected for their robustness, consistently showing *P-values* < 0.05 when analyzed as both continuous and categorical variables. Meanwhile, five out of the 116 SNPs showed significant associations (*P-value* < 0.05) after the FDR adjustment with facial segments in the upper face region using the quasi-landmark mesh approach (Figure 1).

Table 1. SNPs showing significant association with the measurements and indices.

Variable	SNP	*P-value*		Variable	SNP	*P-value*	
Intercanthal width (en'- en')	rs10175706[a]	Continuous	0.009	Superior facial width (fz'- fz')	rs10962767[a]	Continuous	0.007
		2 Categories	0.002			2 Categories	0.001
		3 Categories	0.015			3 Categories	0.008
Biocular width (ex'- ex')	rs10868138[b]	Continuous	0.030	Face width (zy'- zy')	-		
		2 Categories	0.032	Intercanthal index (en'- en' / ex'- ex' x 100)	rs12495832[a]	Continuous	0.012
		3 Categories	0.023			2 Categories	0.001
	rs1454072[c]	Continuous	0.003			3 Categories	0.003
		2 Categories	0.020		rs13097965[b]	Continuous	0.008
		3 Categories	0.015			2 Categories	0.016
Eye fissure width (ex'- en')	rs6129564[c]	Continuous	0.007			3 Categories	0.033
		2 Categories	0.038		rs62578082[a]	Continuous	0.024
		3 Categories	0.038			2 Categories	0.017
	rs62578082[a]	Continuous	0.001			3 Categories	0.038
		2 Categories	0.006	Palpebral fissure length index (en'- ex' / fz'-fz' x 100)	rs12495832[a]	Continuous	0.000
		3 Categories	0.020			2 Categories	0.000
Minimum frontal width (ft'- ft')	rs10175706[a]	Continuous	0.033			3 Categories	0.001
		2 Categories	0.004	Forehead-face width index (ft'- ft' / zy'- zy' x 100)	-		
		3 Categories	0.024				

[a] SNPs revealing previously unreported associations with the upper face, previously linked to the chin [13].

[b] SNPs revealing previously unreported associations with the upper face, previously linked to the nose [13, 17].

[c] SNPs previously associated with the upper face [7, 15].

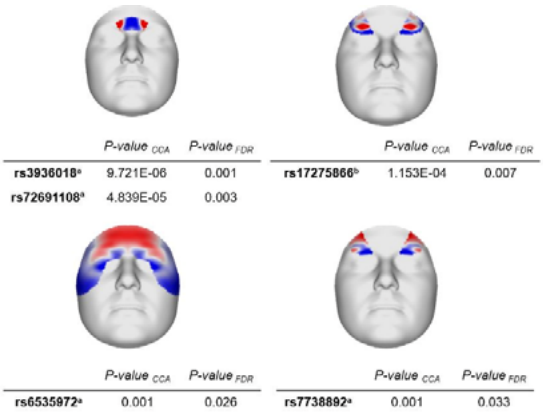


Figure 1. SNPs showing significant association with different facial modules in the upper face. Inward and outward movement are represented in blue and red, respectively. (a)SNPs previously associated with the upper face [4, 7, 16, 18] and (b)SNPs revealing a previously unreported association with the upper face, previously linked to the chin [13].

Discussion

Association analyses in this study revealed significant correlations for several SNPs, although these varied depending on the analytical approach employed.

Using the physical anthropology approach, eight SNPs were found to be significantly correlated with facial measurements and indices in our population (Table 1), with three of these SNPs being associated with more than one facial trait. While some SNPs had previously been linked to the upper face, our results identified novel unreported associations with this region (rs10175706, rs62578082, rs10962767, and rs12495832 were previously related to chin morphology, and rs10868138 and rs13097965 had been related to nasal features in prior studies) [7, 13, 15, 17]. Most of these associations were concentrated in the ocular region, affecting intercanthal, biocular, and eye fissure width, as well as intercanthal and palpebral fissure length indices (Table 1).

In contrast, the quasi-landmark mesh approach identified in our population five SNPs with significant correlations to facial segments in the upper face region (Figure 1). Similar to the physical anthropology method, some SNPs had known associations with the upper face, but our results also uncovered new correlations with this region (rs17275866 was previously related to chin morphology) [4, 7, 13, 16, 18]. The majority of these associations were localized to the forehead area.

The differences in SNPs identified between the two methods likely reflect the distinct facial traits captured by each analytical approach.

Conclusion

This study supports the existence of genetic markers associated with facial morphology in a Spanish population. Of the 13 identified SNPs, six validate correlations with facial regions noted in previous studies, while seven reveal novel associations. In our population, physical anthropology methods demonstrated greater associations with the ocular region, whereas quasi-landmark mesh techniques identified correlations primarily in the forehead region. Therefore, these findings highlight the variability in significant SNPs depending on the analytical approach employed, suggesting that a multifaceted strategy may be necessary for a comprehensive understanding of the genetic basis of facial morphology.

Acknowledgments

Authors acknowledge funding support from Basque Government through predoctoral fellowship PRE_2021_2_0157, as well as project grants IT-1271-19 and IT1633-22.

Conflict of interest statement

Authors declare no competing interest.

References

[1] Kayser M (2015) Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. Forensic Sci Int Genet 18:33–48. https://doi.org/10.1016/j.fsigen.2015.02.003

[2] Adhikari K, Fuentes-Guajardo M, Quinto-Sánchez M, et al (2016) A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. Nat Commun 7:1–11. https://doi.org/10.1038/ncomms11616

[3] Cha S, Lim JE, Park AY, et al (2018) Identification of five novel genetic loci related to facial morphology by genome-wide association studies. BMC Genomics 19:1–17. https://doi.org/10.1186/s12864-018-4865-9

[4] Claes P, Roosenboom J, White JD, et al (2018) Genome-wide mapping of global-to-local genetic effects on human facial shape. Nat Genet 50:414–423. https://doi.org/10.1038/s41588-018-0057-4

[5] Cole JB, Manyama M, Kimwaga E, et al (2016) Genomewide Association Study of African Children Identifies Association of SCHIP1 and PDE8A with Facial Size and Shape. PLoS Genet 12:1–19. https://doi.org/10.1371/journal.pgen.1006174

[6] Crouch DJM, Winney B, Koppen WP, et al (2018) Genetics of the human face: Identification of large-effect single gene variants. Proc Natl Acad Sci USA 115:E676–E685. https://doi.org/10.1073/pnas.1708207114

[7] Hoskens H, Liu D, Naqvi S, et al (2021) 3D facial phenotyping by biometric sibling matching used in contemporary genomic methodologies. PLoS Genet 17:1–28. https://doi.org/10.1371/journal.pgen.1009528

[8] Jacobs LC, Liu F, Bleyen I, et al (2014) Intrinsic and Extrinsic Risk Factors for Sagging Eyelids. JAMA Dermatol 150:836–843. https://doi.org/10.1001/JAMADERMATOL.2014.27

[9] Lee MK, Shaffer JR, Leslie EJ, et al (2017) Genome-wide association study of facial morphology reveals novel associations with FREM1 and PARK2. PLoS One 12:1–13. https://doi.org/10.1371/journal.pone.0176566

[10] Li Y, Zhao W, Li D, et al (2019) EDAR, LYPLAL1, PRDM16, PAX3, DKK1, TNFSF12, CACNA2D3, and SUPT3H gene variants influence facial morphology in a Eurasian population. Hum Genet 138:681–689. https://doi.org/10.1007/s00439-019-02023-7

[11] Liu F, van der Lijn F, Schurmann C, et al (2012) A Genome-Wide Association Study Identifies Five Loci Influencing Facial Morphology in Europeans. PLoS Genet 8: e1002932. https://doi.org/10.1371/journal.pgen.1002932

[12] Paternoster L, Zhurov AI, Toma AM, et al (2012) Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. Am J Hum Genet 90:478–485. https://doi.org/10.1016/j.ajhg.2011.12.021

[13] Pickrell JK, Berisa T, Liu JZ, et al (2016) Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet 48:709–717. https://doi.org/10.1038/ng.3570

[14] Qiao L, Yang Y, Fu P, et al (2018) Genome-wide variants of Eurasian facial shape differentiation and a prospective model of DNA based face prediction. Journal of Genetics and Genomics 45:419–432. https://doi.org/10.1016/j.jgg.2018.07.009

[15] Shaffer JR, Orlova E, Lee MK, et al (2016) Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology. PLoS Genet 12:1–21. https://doi.org/10.1371/journal.pgen.1006149

[16] White JD, Indencleef K, Naqvi S, et al (2021) Insights into the genetic architecture of the human face. Nat Genet 53:45–53. https://doi.org/10.1038/s41588-020-00741-7

[17] Wu W, Zhai G, Xu Z, et al (2019) Whole-exome sequencing identified four loci influencing craniofacial morphology in northern Han Chinese. Hum Genet 138:601–611. https://doi.org/10.1007/s00439-019-02008-6

[18] Xiong Z, Dankova G, Howe LJ, Lee MK (2019) Novel genetic loci affecting facial shape variation in humans. eLife 8:e49898. https://doi.org/10.7554/eLife.49898

[19] Claes P, Walters M, Clement J (2012) Improved facial outcome assessment using a 3D anthropometric mask. Int J Oral Maxillofac Surg 41:324–330. https://doi.org/10.1016/J.IJOM.2011.10.019

[20] Martos R, Valsecchi A, Ibáñez O, Alemán I (2018) Estimation of 2D to 3D dimensions and proportionality indices for facial examination. Forensic Sci Int 287:142–152. https://doi.org/10.1016/J.FORSCIINT.2018.03.037

[21] Heike CL, Upson K, Stuhaug E, Weinberg SM (2010) 3D digital stereophotogrammetry: A practical guide to facial image acquisition. Head Face Med 6:1–11. https://doi.org/10.1186/1746-160X-6-18/FIGURES/7

[22] Skeleton-id. https://skeleton-id.com/.

[23] R Core Team (2021) R: A language and environment for statistical computing

[24] White JD, Ortega-Castrillón A, Matthews H, et al (2019) MeshMonk: Open-source large-scale intensive 3D phenotyping. Sci Rep 9:6085. https://doi.org/10.1038/s41598-019-42533-y

[25] Sero D, Zaidi A, Li J, et al (2019) Facial recognition from DNA using face-to-DNA classifiers. Nat Commun 10:2557. https://doi.org/10.1038/S41467-019-10617-Y

[6] Wilke F, Herrick N, Matthews H, et al (2023) Exploring regional aspects of 3D facial variation within European individuals. Sci Rep 13:3708. https://doi.org/10.1038/s41598-023-30855-x

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 183-189
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Forensic DNA Profiling and Phenotyping from Tobacco Sticks and Nails with Semi-Permanent Polish Samples

Giulia Fazio[1*], Sara Amurri[1], Arianna Giorgetti[1], Filomena Melchionda[2], Chiara Turchi[2], Susi Pelotti[1], Carla Bini[1]

[1]*Department of Medical and Surgical Sciences, Section of Legal Medicine, University of Bologna, Italy.*

[2]*Department of Biomedical Sciences and Public Health, Marche Polytechnic University, Ancona, Italy.*

** Corresponding author at: giulia.fazio2@unibo.it,*

Tel: 0039 051 2088343, University of Bologna, Via Irnerio, 49 Bologna, Italy.

Abstract

The growing popularity of Heat-Not-Burn devices has significantly increased the likelihood of finding tobacco sticks as evidence at crime scenes. Similarly, the increasing global trend in nail beautification may lead to analyse nails with semi-permanent polish for the identification of badly preserved corpses. The aim of this study was to investigate the potential for forensic DNA profiling and phenotyping of eye, hair, and skin colour characteristics using an MPS assay consisting of the 41 SNPs included in the HIrisPlex-S system on 22 samples of which 16 tobacco sticks and 6 clipped fingernails with semi-permanent nail polish. Genotype data obtained was then uploaded on DNA Phenotyping webtool to predict eye, hair, and skin colour. Full STR profiles and phenotype prediction were obtained for all the analysed samples and eye, hair and skin traits of samples donors were predicted with high accuracy.

Keywords

Forensic DNA phenotyping, HIrisPlex-S system, Massively Parallel Sequencing (MPS), Heat-Not-Burn tobacco sticks, Semi-permanent nail polish.

Introduction

In recent years, Forensic DNA phenotyping (FDP) has emerged as a new "DNA intelligence approach". Predicting externally visible characteristics (EVCs) from

DNA can provide investigative leads to identify unknown victims and perpetrators of crimes, as well as disaster victims and missing persons [1]. The widespread diffusion of heat-not-burn (HNB) tobacco products and new nail enhancement techniques increases the chances of finding evidence in crime scene such as the new tobacco sticks and nails embedded with semi-permanent polish useful for human identification [2-3]. The present study aims to explore the possibility of forensic DNA profiling and prediction of eye, hair, and skin colour prediction by FDP analysis of tobacco sticks and clipped fingernails varnished with semi-permanent nail polish.

Material studied, methods, techniques

*A t*otal of 117 samples were collected from 37 Italian volunteers and their written informed consent was obtained. A questionnaire was administered to all volunteers providing basic information on eye, hair and skin colour phenotypes. This study was approved by the Bioethical Commitee of the University of Bologna, Italy (Protocol n.0175936). DNA purification of these samples was performed using the Maxwell® FSC DNA IQ™ Casework Kit (Promega). Quality and quantity of the extracted DNA were determined using the PowerQuant System (Promega) on the QuantStudio 5 Real-Time PCR System for Human Identification (Applied Biosystems), and amplified with GlobalFiler™ IQC PCR Amplification Kit (Applied Biosystems) on the VeritiPro™ Thermal Cycler 96-well (Applied Biosystems) instrument.

Twenty-two out of 117 samples, showing a full single source profile matching the volunteer one, were selected for FDP analysis: 16 tobacco sticks and 6 clipped fingernails with semi-permanent nail polish. The analysis was carried out using an MPS assay consisting of 41 SNPs included in the HIrisPlex-S system [4]. This assay was designed by the Ion AmpliSeq Designer webtool with amplicon sizes <180 bp to allow analysis of degraded DNA [5]. The MPS libraries were performed using Precision ID Library kit (Thermo Fisher Scientific) and sequenced using Ion GeneStudio S5 System (Thermo Fisher Scientific).

The resulting genotype data were then uploaded on the HIrisPlex-S DNA Phenotyping webtool to predict eye (https://hirisplex.erasmusmc.nl/), hair, and skin colour. The eyes and hairs colour of the volunteers were categorised into three groups for eye colour (blue, intermediate, and brown) and four distinct classes for hair colour (red, blond, brown, and black), in accordance with the probability threshold established by Walsh et al. [6]. Moreover, the model proposed by Pospiech et al. [7] was employed for the purpose of evaluating eye colour prediction. Additionally,

with regard to skin colour, the main HIrisPlex-S prediction may be influenced by the second most probable, resulting in a lighter or darker tone than the originally predicted colour. The five skin colour categories predicted by this model are based on the dermatological Fitzpatrick scale for skin colour and sun sensitivity [8].

Results

The average coverage of sequenced libraries was 2511.5 and the uniformity of coverage of 97% (mean). All samples provide complete genotype profiles in all 41 phenotype-informative DNA markers based on the coverage thresholds used.

Accurate predictions of EVCs were achieved for 50% of the 22 sample using the HirisPlex-S system; this percentage increases to 68% when combining the model by Pospiech et al. Detailed results obtained from the HIrisPlex-S Webtool, including the exact prediction probabilities are included in Table 1 (click on the link to view the image). The assay results showed prediction accuracy for the three pigmentation traits, consistent with the performance reported by the original models [6,8]. The prediction accuracy expressed as Area Under the receiver operating characteristic Curve (AUC) values for eye colour were 0.93 for blue, 0.94 for brown, and 0.73 for intermediate. Similarly, the AUC values for hair colour prediction were 0.81 for black, 0.74 for brown, 0.81 for blond, and 0.92 for red. For skin colour prediction were 0.78 for very pale, 0.74 for pale, 0.76 for intermediate, 0.95 for dark, and 0.99 for very dark skin.

Table 1. *Prediction results for 22 samples using the HIrisPlex-S system.*

The prediction success rate was 100% (n = 3/3) for blue eyes and 79% (n = 11/14) for brown eyes. Nevertheless, three samples (14%) of the 22 were inconclusive, with a prediction probability below the established threshold. In 23% (5 out of 22) of cases, the intermediate eye colour phenotype (green) was incorrectly predicted. The majority of these (80%) were predicted as blue, while only one was classified as brown (Figure 1.b). In applying Pospiech's model, it was found that four of the five with incorrectly predicted samples showed a green/blue colour, while one hazel/brown. Furthermore, the three inconclusive samples were predicted as brown, consistent with the volunteers' statement.

A comparison of the predicted results with the volunteers' actual phenotypic data showed that hair colour was correctly predicted in 64% of the samples. The prediction success rates were 100% (n = 4/4) for black hair, 61.5% (n = 8/13) for brown hair, 40% (n = 2/5) for blond hair. No predictions were made for red hair. The most common instances of prediction inconsistency were observed among individuals with blond and brown hair, where darker hair colours, ranging from dark brown to black, were predicted (Figure 1.a). However, in two cases, the incorrect predictions for these categories indicated a lighter hair colour than those reported by the volunteers.

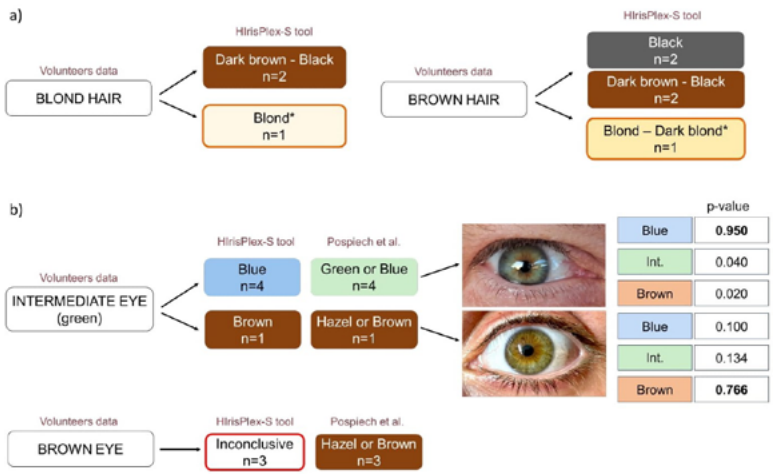


Figure 1. a) *Distribution of prediction inconsistencies for individuals with blond and brown hair. *= two cases reflect incorrect predictions attributed to age-related hair darkening.* **b)** *Eye colour prediction results for intermediate and brown using the HIrisPlex-S and Pospiech et al. models. Volunteer photographs are displayed alongside an example of a discordant prediction. Bold p-values indicate the highest probabilities for the predicted eye colour.*

The prediction was successful for 77% of the samples, while two of the 22 samples yielded inconclusive results. Nine volunteers (41%) exhibited intermediate Fitzpatrick scale categories, falling between II and III/IV (predicted categories: pale/intermediate or intermediate/pale). Two volunteers (9%) were classified as III/IV and V (intermediate or intermediate/dark), and nine (41%) as Fitzpatrick III/IV (intermediate category). In three samples (14%) an intermediate skin colour was incorrectly predicted, despite the samples being classified as very pale.

Discussion

This study highlighted the potential of the HIrisPlex-S system to predict phenotypic traits with high accuracy from DNA samples, such as tobacco sticks and nails with semi-permanent nail polish, demonstrating that these samples are reliable sources of DNA also for FDP analysis. Prediction inconsistencies were observed for more complex phenotypes, such as intermediate eye colour and hair colour ranging from blond to dark brown. This could be attributed to the subjective perceptions of eye and hair colour by the volunteers, the challenges in distinguishing between different shades of hair colour, and the sub-categories of intermediate eye colour shades [8-9]. In hair samples where the predicted colour was lighter than the actual hair colour, further investigation among the volunteers revealed that age-related hair darkening during their childhood [9].

*T*he discrepancies observed in the intermediate eye colour predictions by HIrisPlex-S and the objective data provided by the volunteers, can be clarified by the classification proposed in previous studies [10-11], also supported by photographs collected for the current study. In this way, intermediate eye colour was divided into two categories: light intermediate (often referred to as "green") and dark intermediate (commonly known as "hazel-green"). Indeed, the results showed that the majority of individuals with "light intermediate" eyes were predicted to have blue eyes, while those with "dark intermediate" eyes were predicted to have brown eyes. The eye colour prediction model proposed by Pospiech reduced discordant predictions and produced results that were consistent with the volunteers' statements [8]. In Figure 1.a, an example of a discordant prediction is shown in the case of green eyes, where the HIrisPlex-S model predicted blue eyes. The volunteer photo confirmed a light intermediate eye colour, consistent with the prediction by Pospiech et al. A similar discrepancy was observed in the sample with brown eyes. Therefore, caution should be taken when predicting blue or brown eyes, as individuals may actually have intermediate green or hazel eyes.

As commonly observed in populations of European ancestry [8], the predicted skin colour suggested that the samples analysed in this study belong to the light/intermediate skin category, which darkens or tans with sun exposure during the summer months.

Conclusion

In conclusion, the results showed that the EVCs were correctly predicted in more than half of the samples, tobacco sticks and fingernails varnished with semi-permanent polish, showing consistency with the volunteers' information. Despite challenges in interpreting intermediate eye colour phenotypes, correct predictions were achieved by combining the HIrisPlex-S and Pospiech et al. models. Enhancing the HIrisPlex-S system with additional DNA markers associated with eye color could improve the accuracy of predicting intermediate eye colour categories, which are currently detected with significantly lower accuracy than "blue" and "brown" categories. This approach should reduce potential problems caused by inter-observer perceptual differences that may occur in the practical application.

Acknowledgments

We are grateful to all volunteers who provided biological samples and pigmentation phenotype information to this study. This research was supported by Fondazione del Monte di Bologna e Ravenna and by the Italian Ministry of University and Research for funding the PRIN2022 project, Legal, Ethical and Social Challenges of Forensic DNA Phenotyping in Italy (LetFor) (2022TL8TAE).

Conflict of interest statement

The authors declare that they have no conflict of interest.

References

[1] Kayser M, Branicki W, Parson W, et al., Recent advances in Forensic DNA Phenotyping of appearance, ancestry and age, Forensic Sci Int Genet. 65(2023) 102870.

[2] Putek J, Przybyla T, Szepietowski JC, et al., Side-effects Associated with Gel Nail Polish: A Self-questionnaire Study of 2,118 Respondents. Acta Derm Venereol. 100(2020) adv00325.

[3] Ratajczak A, Jankowski P, Strus P, et al., Heat Not Burn Tobacco Product-A New Global Trend: Impact of Heat-Not-Burn Tobacco Products on Public Health, a Systematic Review, Int J Environ Res Public Health. 17(2020) 409.

[4] Breslin K, Wills B, Ralf A, et al., HIrisPlex-S system for eye, hair, and skin color prediction from DNA: Massively parallel sequencing solutions for two common forensically used platforms, Forensic Sci Int Genet. 43(2019) 102152.

[5] Melchionda F, Silvestrini B, Robino C, et al., Development and Validation of MPS-Based System for Human Appearance Prediction in Challenging Forensic Samples. Genes (Basel). 13(10) (2022) 1688.

[6] Walsh S, Chaitanya L, Clarisse L, et al., Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage, Forensic Sci Int Genet. 9(2014) 150-61.

[7] Pośpiech E, Draus-Barini J, Kupiec T, et al., Gene-gene interactions contribute to eye colour variation in humans, J Hum Genet. 56(2011) 447-55.

[8] Chaitanya L, Breslin K, Zuñiga S, et al., The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation, Forensic Sci Int Genet. 35(2018) 123-135.

[9] Walsh S, Liu F, Wollstein A, et al., The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int Genet. 7(2013) 98-115.

[10] Salvoro C, Faccinetto C, Zucchelli L, et al., Performance of four models for eye color prediction in an Italian population sample. Forensic Sci Int Genet. 40(2019) 192-200.

[11] Sari O I, Simsek SZ, Filoglu G, et al., Predicting Eye and Hair Color in a Turkish Population Using the HIrisPlex System. Genes (Basel). 13(2022) 2094.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 191-197
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Prediction of hair colour from genotype in the Argentine population: a preliminary study

Magalí Nowik[1,2,5], Diana M. Hohl[1,3], María F. García di Marco[4], Rebeca González[1,2,5], María G. García[4], Cecilia I. Catanesi[1,2,5*]

[1]*Laboratorio de Diversidad Genética, Instituto Multidisciplinario de Biología Celular - IMBICE (CONICET-UNLP-CICPBA), La Plata, Argentina.*

[2]*Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, Argentina.*

[3]*Profesional de Apoyo Adjunto (Comisión de Investigaciones Científicas, Provincia de Buenos Aires CIC-PBA), Argentina.*

[4]*Laboratorio MANLAB, Área de Filiaciones, Buenos Aires, Argentina.*

[5]*Consejo Nacional de Investigaciones Científicas y Técnicas CONICET, Argentina.*

**Corresponding author at: IMBICE, 526 e/ 10 y 11 esquina Camino General Belgrano (1900) La Plata, Argentina, tel.: +54-221-4210112, e-mail: ccatanesi@imbice.gov.ar*

Abstract

Hair colour is an external visible phenotypic trait, which is determined by a set of genes that are variable among individuals and among populations. It is possible to predict the physical appearance of an individual with a certain probability, by genotyping a biological sample, being a valuable tool for the identification of persons in the field of forensic genetics.

Based on previous studies of phenotyping, systems for predicting phenotype using genetic data have been developed, such as HIrisPlex-S and Snipper, among others. However, these studies have been carried out mainly on populations of European origin, and may not be directly applicable to admixed populations, as is the case of Argentineans, where genotype-phenotype association could differ, given their multiple ancestry components (Eurasian, Native, and Sub-Saharan). The aim of this work is to describe the genetic variation underlying hair colour in Argentina population. We obtained DNA from 158 samples of mouthwash from donors from the province of Buenos Aires. Five SNPs were genotyped for MC1R gene, rs11547464, rs1805006, rs1805007, rs1805008, rs1805009, and one for SLC45A2 gene rs28777. Samples were visually classified into two colour categories, and on the other hand, using Snipper as a bayesian classifier. The coincidence rate was 56.96%, indicating a correlation between the Snipper results and observations in that percentage of instances, with a higher incidence among individuals with dark hair (58.9% for dark-haired individuals and 41.1% for blondes).

Our results suggest the need for improving the values of prediction of hair colour from genotype information for this admixed population. Given the substantial amount of genetic and phenotypic variability of this population, influenced by various ancestry components, it is crucial to increase the accuracy and applicability in forensic investigations by continuing these studies with more markers and a bigger sample of individuals.

Keywords
External Visible Characteristics, phenotyping, SNPs, Snipper, admixed population.

1. Introduction

Several observable phenotypic characteristics in humans, such as hair colour, eye colour, and skin colour, are known as externally visible characteristics (EVCs), and exhibit considerable variability both among and within populations. These physical traits are influenced not only by environmental factors but also by the expression of specific genes, where various polymorphisms have been identified. Their study provides valuable phenotypic information from biological samples, such as those collected from crime scenes. By analyzing specific Single Nucleotide Polymorphisms (SNPs), it is possible to predict with certain probability the physical appearance of an individual, which can be particularly useful in forensic contexts for identifying missing persons or solving criminal cases [1,2,3].

In forensic genetics, phenotyping systems that predict with certain accuracy hair, eye, and skin colour from genetic data have been developed, including tools such as HIrisPlex-S [1,4] and Snipper [5]. Although these systems can provide crucial information in forensic investigations, they have been validated predominantly on populations of European origin. As such, their applicability to admixed populations with diverse genetic backgrounds, such as Argentineans—who have a particular admixture of Eurasian, Native American, and Sub-Saharan ancestries—may be limited, as genotype-phenotype associations could differ significantly [6,7,8].

The aim of this study is to describe the genetic variation underlying hair colour in the Argentine population, contributing to the development of more accurate phenotyping tools for use in forensic science within diverse genetic contexts.

2. Materials and methods

2.1. DNA samples

DNA was obtained from 158 donors from the province of Buenos Aires (114 females and 44 males) aged between 18 and 55 years old. The project was approved

by the Ethics Committee for Biomedical Research at IMBICE, and donors provided individual written informed consent. A buccal rinse with 2 ml of water was collected to obtain exfoliated cells from the buccal mucosa, mixed with 2 volumes of 96% ethanol, and stored at -20°C until processing. DNA extraction followed a standard protocol which involved lysis with Proteinase K and extraction using lithium chloride and chloroform-isoamyl alcohol [9].

2.2. Genotyping and hair colour classification

Specific oligonucleotides were designed for this work, and the target regions were PCR-amplified. The polymorphisms (Table 1) were resolved either through allele-specific PCR and 2% agarose gel electrophoresis for 1 SNP of *SLC45A2* gene, or through Sanger sequencing for 5 SNPs of *MC1R* gene.

Hair colour was determined through comparison by three independent observers, grouped into two main categories: light/blonde and dark/black.

Additionally, hair colour was predicted from genotype using Snipper as a Bayesian classifier.

Table 1. SNPs analyzed in this work; a: GRCh38 38.1/141

SNP	Gene	Chromosome	Position[a]
rs11547464	*MC1R*	16	89919683
rs1805006	*MC1R*	16	89919510
rs1805007	*MC1R*	16	89919709
rs1805008	*MC1R*	16	89919736
rs1805009	*MC1R/TUBB3*	16	89920138
rs28777	*SLC45A2*	5	33958854

2.3. Statistical analysis

Hardy-Weinberg equilibrium (HWE) was calculated with Arlequin v3.5.2.2 [10]. The FST values were calculated with GenAlEx v6.5 [11] for measuring the genetic differentiation between pairs of populations. Interpopulation comparisons were performed using data from 1000 Genomes [12] for African, American, and European populations (total from each continent). SNPStats [13] provided additional statistical tools for analyzing SNP data and evaluating the association of genetic variants with phenotypes [13].

4. Results

The coincidence rate between the Snipper results and the observational data was 56.96%, suggesting a possible correlation between prediction and visual assessment. This correlation was more pronounced among individuals with dark/black hair, for which the coincidence rate was 58.9%, compared to 41.1% for individuals with light/blonde hair. This differential incidence could indicate varying genetic influences or environmental factors affecting hair pigmentation across different groups. Table 2 illustrates the genotype frequencies observed in this study; they fit HWE except for rs1805009.

Table 2. Observed genotype frequencies for the analyzed SNPs

SNP	Genotypes	Frequencies
rs28777	AA	0.46
	AC	0.37
	CC	0.17
rs11547464	AG	0.02
	GG	0.98
rs1805008	CC	0.96
	CT	0.03
	TT	0.01
rs1805007	CC	0.97
	CG	0.01
	CT	0.02
rs1805009	AG	0.01
	CC	0.04
	CG	0.04
	GG	0.91
rs1805006	CC	0.99
	CT	0.01

The FST values revealed significant genetic differentiation between the Buenos Aires population and other populations, including African (FST=0.35), American (FST=0.04), and European (FST=0.18) groups. All observed FST values were statistically significant (p-values<0.01).

5. Discussion

SNP-based tools can provide valuable forensic insights, but their accuracy is limited when they are applied to highly admixed populations. Our results showed a moderate prediction accuracy for hair colour, particularly among dark-haired individuals. This finding is consistent with previous research indicating that systems like Snipper and HIrisPlex, which are optimized for European ancestry, may struggle with the genetic complexity of admixed populations [7,14]. In a Brazilian study, for instance, the HIrisPlex system showed reduced sensitivity in predicting intermediate phenotypes for individuals with mixed European, African, and Native American ancestry [15].

The FST values further contextualize these findings. Specifically, the highest differentiation found between Buenos Aires and African populations suggests a limited gene flow between them, reflecting historical or evolutionary barriers to genetic exchange. In contrast, the moderate genetic differentiation from Europeans indicates a more considerable level of genetic exchange and interaction between Buenos Aires and European populations. Moreover, the differentiation between Buenos Aires and Native American populations was the lowest - although significant - suggesting a relatively higher degree of genetic similarity and gene flow between them. Significant FST values affirm the robustness of these findings and the presence of distinct genetic differentiation patterns between the Buenos Aires population and the other studied populations.

By examining genetic differentiation across diverse populations, we can better understand how historical gene flow has shaped the genetic landscape of Buenos Aires.

6. Conclusion

Our study gives evidence of the challenges of using forensic genetic tools developed for European populations when they are applied to admixed populations. The distinct genetic differentiation observed for the Buenos Aires population impacts the tool´s ability to accurately predict traits of complex genetic background [14], and it further underscores the need for refining prediction models.

It is crucial to increase the accuracy and applicability of phenotyping tools in forensic investigations by continuing these studies with more markers and a larger sample of individuals, thus helping to avoid the potential overlooking of unique local adaptations or unknown variations.

7. Acknowledgments

This work was funded by grants from CONICET, Argentina (PIP 2021 N°2111), FONCyT-MinCyT, Argentina (Préstamo BID PICT 2020 N°01075), and UNLP, Argentina (Proyecto I+D 2024-2025 UNLP - FCNyM N°1018).

We wish to thank all volunteers who participated in this study.

8. Conflict of interest statement

No Conflict of interest.

9. References

[1] L. Chaitanya, K. Breslin, S. Zuñiga et al., The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation, Forensic Sci. Int. Genet. 35 (2018) 123–135.

[2] M. Kayser, Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes, Forensic Sci. Int. Genet. 18 (2015) 33–48.

[3] O. Maroñas, J. Söchtig, Y. Ruiz et al., The genetics of skin, hair, and eye color variation and its relevance to forensic pigmentation predictive tests, Forensic Sci. Rev. 27(1) (2015) 13–40.

[4] S. Walsh, L. Chaitanya, L. Clarisse et al., Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage, Forensic Sci. Int. Genet. 9 (2014) 150–161.

[5] Y. Ruiz, C. Phillips, A. Gomez-Tato et al., Further development of forensic eye color predictive tests, Forensic Sci. Int. Genet. 7(1) (2013) 28–40.

[6] M. Muzzio, J.M.B. Motti, P.B. Paz Sepulveda et al., Population structure in Argentina, PLoS One 13(5) (2018) e0196325.

[7] D.M. Hohl, R. González, G.P. Di Santo Meztler et al., Applicability of the IrisPlex system for eye color prediction in an admixed population from Argentina, Ann Hum Genet 86(6) (2022) 297-327.

[8] H. Dopazo, A.S. Llera, M. Berenstein et al., Genomas, enfermedades y medicina de precisión: un Proyecto Nacional, Ciencia, Tecnología y Política 2(2) (2019) 1–10.

[9] N.J. Gemmell, S. Akiyama, An efficient method for the extraction of DNA from vertebrate tissues, Trends Genet. 12(9) (1996) 338–339.

[10] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10(3) (2010) 564–567.

[11] R. Peakall, P.E. Smouse, GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update, Bioinformatics 28(19) (2012) 2537–2539.

[12] The 1000 Genomes Project Consortium, A global reference for human genetic variation, Nature 526 (2015) 68–74.

[13] X. Solé, E. Guinó, J. Valls et al., SNPStats: a web tool for the analysis of association studies, Bioinformatics 22(15) (2006) 1928–1929.

[14] A. Freire-Aradas, Y. Ruiz, C. Phillips et al., Exploring iris colour prediction and ancestry inference in admixed populations of South America, Forensic Sci. Int. Genet. 13 (2014) 3–9.

[15] T.M. Telles Carratto, L. Marcorin, G. Valle-Silva et al., Prediction of eye and hair pigmentation phenotypes using the HIrisPlex system in a Brazilian admixed population sample, Int. J. Leg. Med. 135 (2021) 1329–1339.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 199-208
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Highly Portable, Specific and Sensitive Platform for Fast-Track Wildlife Species Identification

Nidhi Subhashini[1*], Alexander Griberman[1], Walter Stühmer[2], Daniel Vanek[3,4,5,6]

[1] *SERATEC Gesellschaft für Biotechnologie mbH, Goettingen, Germany.*

[2] *Max Planck Institute of Multidisciplinary Sciences, Goettingen, Germany.*

[3] *Bulovka University Hospital, Prague, Czech Republic.*

[4] *Institute for Environmental Sciences, Charles University, Prague, Czech Republic*

[5] *Forensic DNA Service, Prague, Czech Republic.*

[6] *Department of Forensic Medicine, Second Faculty of Medicine, Charles University, Prague, Czech Republic.*

** Corresponding author at: Ernst-Ruhstrat Str. 5, 37079, Goettingen, Germany, nidhi.subhashini@seratec.com*

Abstract

Rapid and accurate wildlife species identification is essential for biodiversity conservation, ecological research, and wildlife protection enforcement. Traditional identification methods are slow, require expert knowledge, and can lead to misidentification. This study introduces a novel, portable, and highly sensitive platform designed for fast species identification in field settings. Combining advanced molecular biology techniques with compact, low-power hardware, the system can identify a wide range of species using minimally invasive samples like hair, faeces, or blood. It uses quantitative isothermal amplification for highly sensitive and species-specific detection. Results are generated in under an hour, offering near real-time feedback. Initial validation shows high accuracy, comparable to traditional methods. The platform's portability and ease of use make it a scalable and accessible tool for conservation efforts, addressing challenges like poaching and habitat loss in remote or biodiverse regions.

Keywords

Wildlife Forensics, Remote species identification, Wildlife trafficking.

Introduction

Wildlife preservation is crucial for ecological balance, climate regulation, and sustainable livelihood supporting human well-being and evolution. Wildlife conservation contributes to economic prosperity and holds cultural and spiritual importance, preserving traditions and connections to the natural world, particularly for indigenous communities. Many view wildlife preservation as a moral obligation, emphasizing the intrinsic value of all living organisms and the need to minimize human-induced suffering and extinction. However, species extinction due to the illegal trafficking of endangered animals and animal products is a significant issue worldwide. It has been recognized as one of the UN Sustainable Development Goals. Wildlife trafficking poses numerous challenges due to its intricate and diverse nature since trafficking networks span international borders [1, 2], complicating law enforcement's efforts to combat them due to varying legal systems and enforcement capacities among countries. The trade is highly lucrative, incentivizing traffickers to persist despite the risks, driven by the demand for rare species and their parts [3]. In addition, many countries need more resources for effective enforcement and conservation efforts, hindering habitat protection and patrols. Hence, a comprehensive approach involving collaboration among governments, law enforcement, conservation groups, communities, and international stakeholders is crucial. This approach should focus on enhancing enforcement capacity, reducing demand through education and sustainable alternatives, strengthening legal frameworks, and addressing underlying issues like poverty and corruption.

One of the significant obstacle in combating illegal wildlife trade is animal species identification. Testing field wildlife species presents unique challenges compared to laboratory animals or humans (Table 1). High diversity of species, complex morphological variations, processed animal tissues, evasive tactics by traffickers added with technological and methodological limitations complicate the process of correct animal identification. While technologies like DNA barcoding, forensic analysis and high resolution imaging exist, they are not always available or used uniformly across different regions and enforcement bodies. Overcoming these challenges demands interdisciplinary collaboration and technological advancements like portability, remote-sensing and non-invasive sampling methods to enhance data accuracy and reliability in field wildlife testing [4, 5]. Advances in isothermal amplification technologies have led to development of portable devices that can potentially be adapted by law enforcement agencies or customs officials to rapidly identify trafficked wildlife species at various checkpoints. Isothermal

Amplification-based assays can amplify DNA rapidly and efficiently at constant temperature. These assays are simple to perform making them suitable for use in field conditions or resource-limited settings while still providing with high confidence results. This study introduces a portable and highly sensitive platform for rapid wildlife species identification using minimal biological samples. It provides a scalable tool for conservation and law enforcement by enabling on-site DNA analysis with near real-time results, facilitating efforts in combating wildlife trafficking and habitat loss.

Materials and Methods

A step-by-step flowchart showing the Loop-Mediated Isothermal Amplification (LAMP) process, from sample collection (e.g., hair, blood, tissue) to DNA extraction, amplification, and detection is depicted in Figure 1A.

SaLux19 Device

The SaLux19 device is a portable, real-time DNA analysis tool used in wildlife and forensic sciences (Figure 1B). Utilizing loop-mediated isothermal amplification (LAMP), it amplifies specific DNA sequences at a constant temperature of 64°C without the need for complex thermocycling equipment. Its small size (14.5 cm × 9.5 cm × 4 cm) and user-friendly interface make it ideal for on-site testing. Equipped with fluorescent detection, the device provides rapid and sensitive identification of DNA from biological specimens, such as *Sus scrofa* (wild boar), within 40 minutes, offering both qualitative and quantitative analysis [5].

Reagents

Quick-DNA Miniprep Kit (cat. no. D3024, ZymoResearch, Irvine, CA, USA) was used for DNA extraction. iQ Sybr Green Supermix (cat. no. 1708880, Bio-Rad, Hercules, CA, USA) was employed for quantifying nuclear DNA (nDNA). TaqMan™ Control Genomic DNA (cat. no. 4312660, Thermo Fisher Scientific, Waltham, MA, USA) was used as a human DNA control. SYBR Green Gel Stain (cat. no. S7563, Thermo Fisher Scientific, Waltham, MA, USA) was used for fluorescence detection in the LAMP assay. ZebraShield (cat. no. ZEB-S10, SERATEC, Göttingen, Germany) was used to preserve the biological samples during storage and further processing, ensuring sample integrity. Nuclease-Free Water (Thermo Fisher Scientific, Waltham, MA, USA) was used as a solvent in all reaction mixtures. DipStick DNA

Extraction Kit (Bento Bioworks Ltd., London, UK) was employed for alternative DNA extraction via direct lysis. All reagents were prepared and used following the manufacturer's guidelines.

Method

Biological samples from wildlife, such as tissue, blood, or other fluids were collected. For forensic and conservation purposes, samples from endangered species can also be obtained from non-invasive sources like hair or scat. Samples were stored at appropriate temperatures and buffer solutions (ZebraShield) to preserve DNA integrity until analysis.

DNA extraction using a kit was optimized for wildlife samples (e.g., Quick-DNA Miniprep Kit, ZymoResearch) according to the manufacturer's instructions to ensure high-quality DNA suitable for amplification. Elution was performed in 50 µL of nuclease-free water.

For primer-design species-specific genes (e.g., mitochondrial genes like cytochrome c oxidase I [COI] for species-level identification) were identified. Several primer sequences for LAMP assay were designed using bioinformatics tools (e.g., NEB LAMP Primer Design Tool) based on the target species' genetic sequences available in public databases (NCBI, GenBank) or adopted from literature.

The LAMP reaction mixture was prepared with DNA template (from extracted samples), primers specific to the target species LAMP reagents (e.g., Bst polymerase, dNTPs, buffer) and fluorescent dye for real-time detection (e.g., SYBR Green). LAMP assay was performed according to the New England Biolabs (NEB, Ipswich, MA, USA) LAMP protocol. Reaction was performed at 64°C for 30-40 minutes using the SaLux19 device, which allows real-time detection of fluorescence.

Results

The SaLux19 device measured the increase in fluorescence after every amplification cycle during amplification period. The amount of fluorescence correlated with the presence of species-specific DNA in the sample. Positive detection was indicated by a significant increase in fluorescence, and results were available within 40 minutes. The device displayed data analogous to Ct values used in qPCR.

Cross-validations were performed using DNA from closely related species to test the specificity of the primers and amplification. DNA from multiple species were tested ensuring the assay only amplifies the target species' DNA. No

cross-reactivity occurred with non-target species. The output from the SaLux19 can be visualized in real-time on the device's display, showing the amplification curve (Figure 1C). A report of species identification was generated, which includes Ct-like values from the amplification data and a confirmation of the species present in the sample (e.g. Figure 1D). Successful species identification of *Sus scrofa* using this platform and methodology was validated and published by Ruszova et al., 2024 [5].

Discussion

The development of a rapid and portable DNA analysis system using the SaLux19 device leverages Loop-Mediated Isothermal Amplification (LAMP) technology for species identification. This advancement has significant implications for wildlife law enforcement efforts, particularly in the fight against poaching, illegal wildlife trade, and food fraud involving animal products.

LAMP technology allows for the rapid, sensitive, and specific identification of biological specimens even in minute quantities. The SaLux19 device's portability and real-time DNA amplification capabilities make it a valuable tool for wildlife conservation agencies. Traditional species identification methods, such as Polymerase Chain Reaction (PCR), require laboratory settings and complex equipment, often leading to delays. In contrast, the SaLux19-based LAMP assay enables on-site species verification in less than 40 minutes, which can expedite enforcement actions at crime scenes, such as border checkpoints or markets.

By enabling faster and more accurate species identification, this technology strengthens law enforcement's ability to detect and intercept illegal wildlife products. For instance, it allows the identification of endangered species in food products or seized animal parts, such as tusks or hides, helping ensure compliance with regulations like the Convention on International Trade in Endangered Species (CITES). Furthermore, the device's ability to detect Sus scrofa (wild boar) nucleic acids suggests it could be adapted for various species commonly targeted in wildlife crimes.

Altogether, the development of such portable DNA analysis tools enhances the capability of law enforcement agencies to protect biodiversity by providing a reliable means of identifying species in the field, thereby supporting more timely and effective prosecution of wildlife-related offenses (Figure 2).

Conclusion

The development and validation of the SaLux19 device using loop-mediated isothermal amplification (LAMP) offers a significant advancement in remote, portable wildlife species identification. The system's compact design, rapid results, and capability for on-site testing make it a valuable tool for wildlife forensics, particularly in resource-constrained environments. Its real-time monitoring and dual detection methods ensure reliable and efficient identification, meeting the growing need for mobile, cost-effective forensic tools (Table 2). This innovation has potential applications in species conservation, wildlife crime investigation, and expanding forensic methodologies.

Acknowledgments

The authors are sincerely thankful to the International Society for Forensic Genetics (ISFG) for organizing valuable conference, ISFG 2024, and providing the opportunity to present this research.

Conflict of interest statement

The authors declare no conflict of interest.

References

[1] TRAFFIC, WWF India. (2023). India's Pangolins buried in Illegal Wildlife Trade.

[2] Kathy Rock, Jon Garcia. (2024). An Overview of Seizures of CITES-listed Wildlife in the EU in 2022.

[3] Oliver Wright, Simone Louw, Nicola Okes, Alexis Breun (2024). The Online Trade in high-value marine products from South Africa. TRAFFIC.

[4] Hebenstreitova, K., Salaba, O., Trubac, J., Kufnerova, J. & Vanek, D. (2024). The Influence of Tanning Chemical Agents on DNA Degradation: A Robust Procedure for the Analysis of Tanned Animal Hide-A Pilot Study. Life (Basel). 14. https://doi.org/10.3390/life14010147.

[5] Ruszova, E., Vanek, D., Stuhmer, W., Khaznadar, Z. & Subhashini, N. (2024). The Utilization of the SaLux19-Based Loop-Mediated Isothermal Amplification

(LAMP) Assay for the Rapid and Sensitive Identification of Minute Amounts of a Biological Specimen. Life (Basel). 14. https://doi.org/10.3390/life14050579.

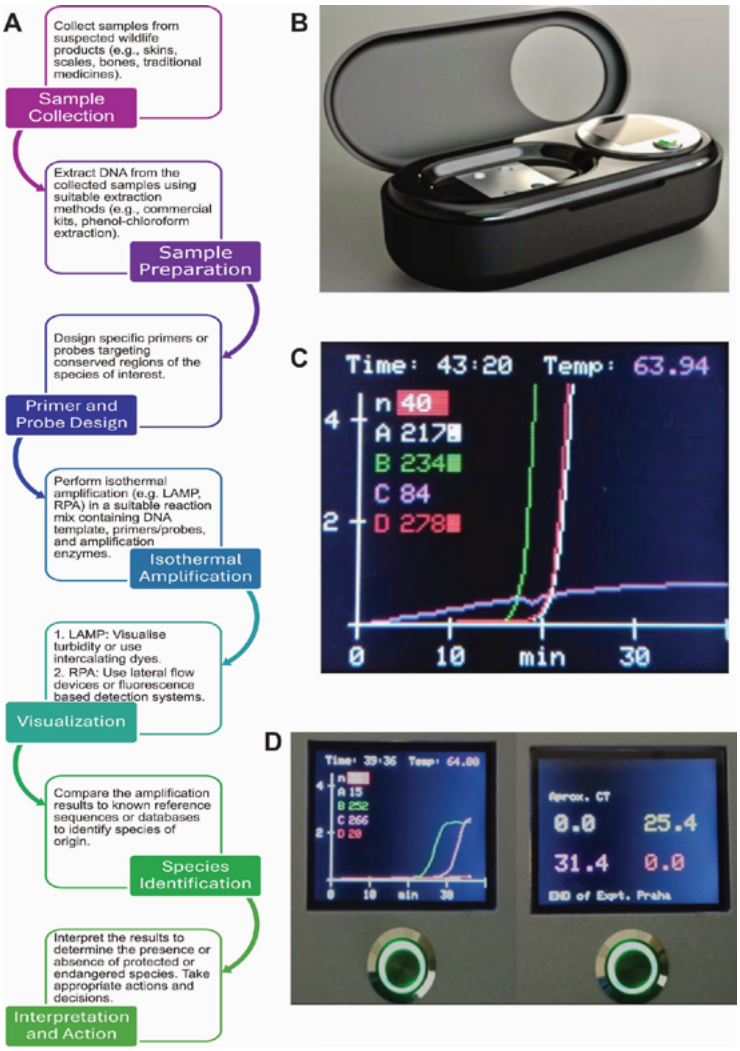


Figure 1. A: LAMP Assay Workflow; B: SaLux19 portable LAMP device; C: A real-time graph demonstrating the amplification curve, where the increase in fluorescence over time is plotted, illustrating how positive species detection is identified within the 40-minute window; D: The device displayed data analogous to Ct values used in qPCR.

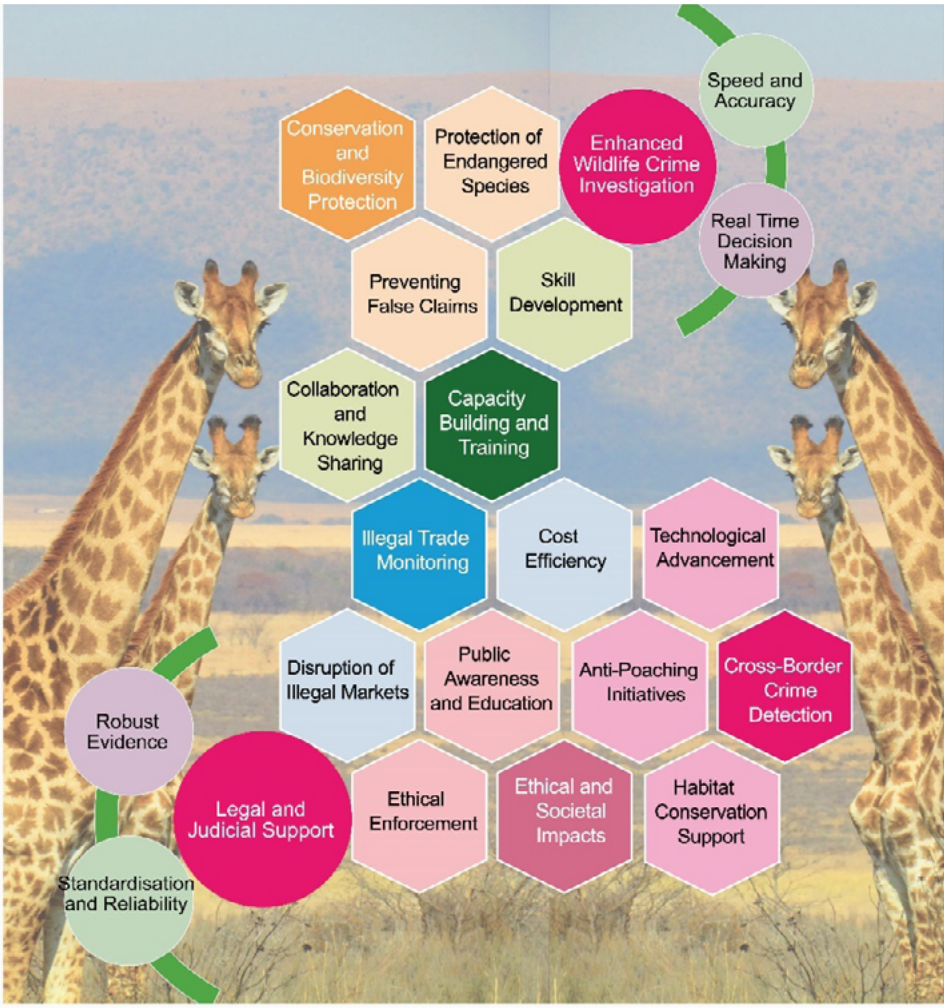


Figure 2. Impact of the rapid and efficient animal species identification assay developed on the portable SaLux19 device.

Table 1: The unique challenges of wildlife species identification compared to lab animals or humans.

Challenge	Wildlife Species	Lab Animals	Humans
Genetic Diversity	High genetic diversity within species.	Limited genetic diversity (controlled breeding).	Moderate genetic diversity with known genetic records.
Environmental Impact	Varying environments can alter appearance and behaviour.	Controlled environments, minimal variation.	Standardized living conditions, minimal influence on identification.
Access and Observation	Difficult to access in natural habitats, limited observation time.	Easily accessible in controlled labs.	Readily available for testing and observation.
Data Collection	Often relies on indirect evidence (e.g., tracks, camera traps).	Direct data collection is possible (behaviour, DNA).	Extensive and detailed data collection available (medical history, biometrics).
Physical Variation	High intraspecies variation (e.g., size, coloration, sexual dimorphism).	Low variation due to controlled breeding.	Physical variation is well-documented (e.g., height, weight, facial features).
Species Misidentification	Many similar species, leading to frequent misidentification.	Limited species and well-documented phenotypes.	Clear distinctions between individuals and populations.
Genomic Reference Databases	Limited reference genomes for many species.	Well-established reference genomes.	Comprehensive reference genomes available (e.g., Human Genome Project).
Non-invasive Identification Techniques	Challenging to implement; reliance on remote techniques like camera traps, drones, DNA from environmental samples.	Easier to implement due to controlled environment and handling.	Commonly available techniques like fingerprinting, facial recognition, and DNA testing.
Ethical Considerations	Impact on species and ecosystem needs consideration (endangered species, stress).	Regulated by animal welfare standards.	Strict ethical standards for research involving humans.
Behavioural Data	Difficult to gather consistent behavioural data across wild populations.	Consistent behavioural data due to controlled environment.	Behavioural data is highly documented (psychological tests, interviews).
Technological Limitations	Use of technology (e.g., satellite, bioacoustics) is often expensive and prone to errors.	Basic technologies (RFID chips, lab-based imaging).	Advanced technologies for biometric and genetic identification are widely available.
Taxonomy and Classification	Ongoing revisions in taxonomy (new species, reclassification).	Well-established species taxonomy.	Fixed classification systems (e.g., ethnicity, nationality).
Sampling Techniques	Often invasive or challenging (e.g., tranquilization, tissue sampling).	Easier sampling due to lab conditions.	Non-invasive techniques are the norm (blood tests, saliva).

Table 2: The major advantages of the SaLux19 LAMP technology over traditional PCR-based methods, particularly for field-based wildlife identification and conservation efforts.

Feature	Traditional PCR-Based Methods	SaLux19 LAMP Technology
Portability	Requires large, stationary laboratory equipment (e.g., thermocycler).	Highly portable, compact device (14.5 cm × 9.5 cm × 4 cm), ideal for fieldwork.
Speed	Takes several hours (including sample preparation, amplification, and analysis).	Results available in less than 40 minutes.
Sensitivity	High sensitivity, but may require multiple steps like DNA extraction and optimization for high-quality samples.	Comparable sensitivity, detects minute quantities of DNA with minimal sample preparation.
Equipment Complexity	Requires complex thermocycling equipment and multiple reagents (often run by trained technicians).	Simple setup with isothermal amplification, no need for complex thermocycling; user-friendly interface.
Sample Types	Can handle a wide range of sample types but often requires high-quality samples and lab-grade processing.	Can process minimally invasive samples (hair, faeces, blood) with on-site DNA extraction.
Real-Time Monitoring	Requires additional equipment (e.g., real-time PCR machine) for real-time monitoring of DNA amplification.	Real-time fluorescence detection is built into the SaLux19 device.
Power Requirements	Requires stable power supply (lab-based electricity) to run the thermocycler and other equipment.	Low-power consumption, can operate with portable battery packs in field settings.
Use Case	Best suited for controlled laboratory environments.	Optimized for remote fieldwork and on-site species identification.
Cost Efficiency	Expensive due to equipment, reagents, and infrastructure needs.	More cost-effective, requiring minimal equipment and reagents, suited for resource-limited settings.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 209-215
DOI: https://dx.doi.org/10.15304/cc.2025.1869

The Canine DNA Recovery Project: Current Findings and Next Steps

Louise Dawnay[1], Paul Riley[1], Matt Lewis[1], Suzzanne McColl[1,], Nick Dawnay[1*]

[1] *School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University*

** Corresponding author at: email: n.m.dawnay@ljmu.ac.uk, tel: +44 151 2312485*

Abstract

The Canine DNA Recovery Project (CDnaRP) is a collaborative forensic project that aims to develop best practice methods for the collection and analysis of dog DNA recovered from attacked livestock and wildlife. The project works closely with police and forensic practitioners to optimise and standardise methods and also works with special interest groups including vets, rural insurers, farmers, and charities to ensure stakeholder awareness of the research and how it relates to livestock and wildlife offences involving canines. Our research to date has developed and validated a robust qPCR assay for canine DNA quantification. The method acts as an important quality control step in the forensic DNA analysis pipeline and has allowed us to assess the effectiveness of different DNA recovery techniques such as swabbing, taping, and cutting. Other areas of our research have revealed the extent to which PCR inhibitors and livestock DNA prevent PCR amplification and canine DNA profile interpretation and has identified mitigation steps leading to improved data quality. To build on these findings, our project has developed an Early Evidence Kit (EEK) for police and rural stakeholder groups to use in the event of a livestock attack. The distribution of these kits will allow a greater number of samples to be collected for the research and will begin to understand whether non-enforcement groups such as farmers themselves can collect admissible evidence in the future. Our presentation looks at the validation of this kit and considers their application in the UK Criminal Justice System.

Keywords

Canine DNA, qPCR, Sample Recovery, Crime Scene.

Introduction

In the United Kingdom, livestock attacks by dogs are a 'non-recordable offence', meaning police are not required to record their frequency. The most comprehensive

police data to date recorded 1705 incidents between 2013 and 2017, which resulted in 1928 livestock killed and 1614 livestock injured [1]. This is likely an underestimate as many incidents go unreported to police and data from insurance claims suggests thousands of attacks occur each year, costing £2.4 million in 2023 alone [2]. Reducing the frequency of dog attacks on livestock is a rural policing concern and is managed through initiatives including educational training programs [3], targeted police operations [4] and proposed changes to UK legislation [5]. Forensic support can aid police investigations by establishing a link between the suspect dog and attacked livestock through canine STR profiling, but this approach is not often used. Discussion with stakeholders identified issues within the sample collection and testing pipeline and led to the creation of the Canine DNA Recovery Project (CDnaRP), a multiphase, multi-stakeholder project led by Liverpool John Moores University, which aims to develop, promote, and apply best practice methods for the collection and analysis of canine DNA from attacked livestock and wildlife.

Methods

a. Community Building and CDnaRP Structure

The UK contains many stakeholders concerned with wildlife and rural crime including government, private companies, charities, professional societies, and academic researchers. The CDnaRP works with these groups to understand their needs, with stakeholders sitting on the project Scientific Advisory Board (Figure 1a).

b. Identifying Knowledge and Research Gaps

Discussion with police and forensic providers highlighted that the canine forensic DNA testing pipeline is underdeveloped and poorly characterised compared to that used in human forensic analysis (Figure 1b). Currently, there is no method to localize canine saliva on submitted swabs or samples, and only a single method (swabbing) recommended to recover DNA from attacked animals [6] despite human research suggesting other methods may offer advantages [7]. Issues with PCR inhibition during STR amplification have not been assessed despite evidence that many submitted samples fail to amplify during PCR. Finally, a standardized method for STR profiling for use in livestock attacks is needed, including robust STR chemistry that minimizes PCR inhibition and livestock amplification. Such laboratory improvements have to be balanced against the fact that police are often unable to collect a forensic sample due to limited resources and a statutory duty for the livestock keeper to dispose of the animal carcass [8].

Results and Discussion

a. Optimisation of recovery methods

A robust qPCR assay was developed to allow accurate quantification of recovered canine DNA [9]. The developed triplex assay amplifies MC1R for autosomal detection (Figure 1c), SRY for male canine detection (Figure 1d), and a synthetic IPC to detect the presence of inhibitors (Figure 1e). The assay was suitably sensitive and specific, and passed all common validation studies [10].

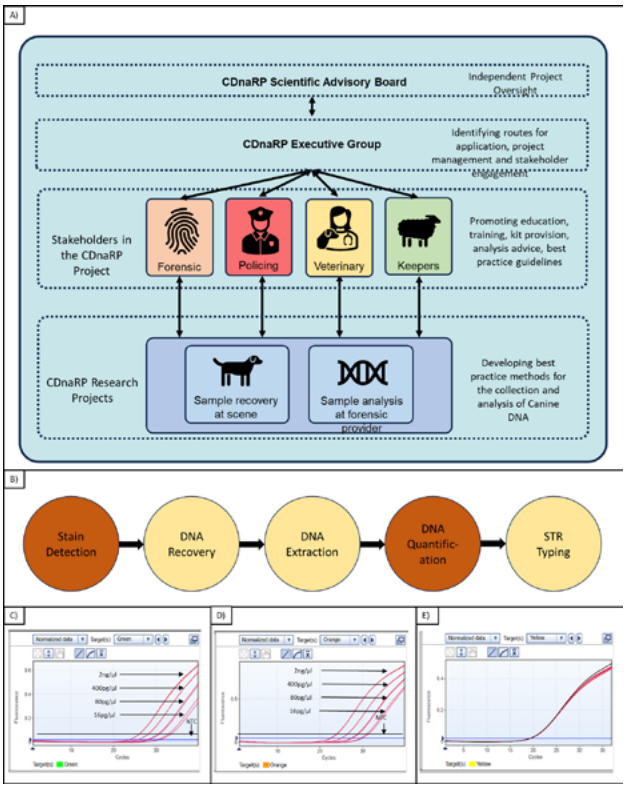


Figure 1. A) Structure of Canine DNA Recovery Project and its remit; B) Canine DNA Pipeline showing areas for development. Red = no current standardised method; Yellow = method needs assessment. C) qPCR amplification plot of MC1R; D) qPCR amplification plot of SRY; E) qPCR amplification plot of Internal positive control.

The qPCR assay was used to compare the amount of canine DNA recovered using three methods; a) swabbing, b) mini-taping, c) cutting wool. Recovery from

naturally shed wool spiked with canine saliva showed swabbing was the least effective method with significantly greater canine DNA recovery using scissors (Figure 2a). The same trend was observed when these methods were applied to real-world samples collected from livestock attacks in Spring 2023, although there was a ~20-fold decrease in DNA yield (Figure 2b) suggesting mock samples were over-spiked. Results also revealed that average amounts of DNA recovered by swabbing (~20pg/µl) was insufficient to generate a STR profile and provides one explanation for the observed analytical pipeline failures.

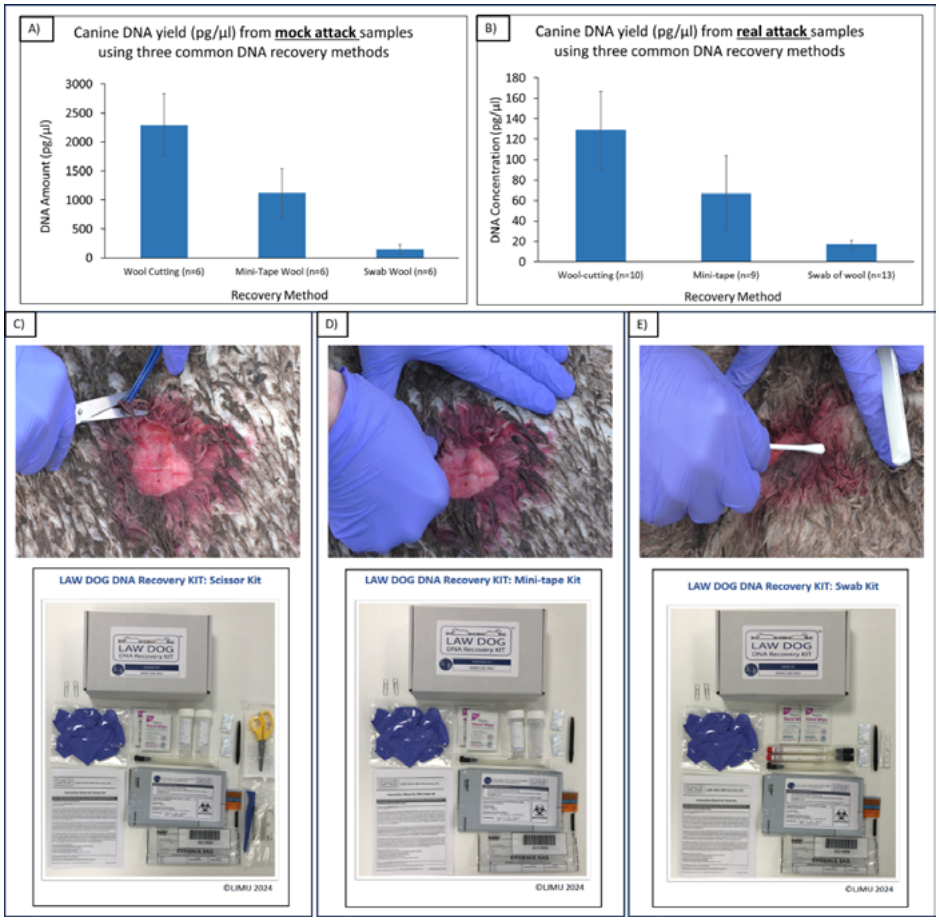


Figure 2. A) Amount of canine DNA (pg/µl) recovered from mock attack samples; B) Amount of canine DNA (pg/µl) recovered from real attack samples. Three different recovery methods and their associated materials used in collection; C) Scissor Kit, D) Mini-Tape Kit, E) Swab Kit.

Data also revealed instances where qPCR detected canine DNA but STR typing subsequently failed due to PCR inhibition. Assessment of methods to combat PCR inhibition, including different DNA extraction kits, sample dilution, and STR formulations and panels, is currently being written up as a technical note. Results of this research also found substantial amplification of livestock DNA using the Canine Genotype 2.1 Kit [11], which was much reduced when using CADNAP Panels 1 and 2 [10, 12], leading to an ongoing collaborative effort to develop a standardised approach for analysis of mixed canine:livestock samples. Guidelines are being developed for both Canine Genotype 2.1 Kit and CADNAP Panels 1 and 2 to allow analysis flexibility for forensic laboratories. Together these data support changes to the existing DNA recovery method and analytical pipeline used by UK police and forensic providers in response to livestock attacks.

b. Widening participation through citizen science

An ambitious citizen science project has been launched to further assess and promote the developed sample collection methods by providing free training and collection kits to police, vets and livestock keepers. This will i) assess the use of methods across a wider group of individuals, ii) derive DNA data to augment existing data and assess reproducibility, and iii) serve to validate the methods and end-user groups.

The developed LAW (Livestock And Wildlife) DOG DNA Recovery Kits contain everything needed to successfully recover, preserve, and secure canine biological evidence at the scene of a livestock attack (Figure 2c-e). Kit release in Summer 2024 was well received by stakeholders and the rural community, resulting in positive press attention [25, 26]. KA future possibility is where the kits are used beyond the scope of the research project, with police, vets and livestock keepers undergoing certified training, to collect biological samples alongside digital data to submit for forensic analysis. This would require acceptance by the UK criminal justice community with the data collected being admissible in court on a case-by-case basis. While this may necessitate a change to current forensic regulations, it would solve issues associated with police response times, meaning evidence was collected early by the livestock keeper and/or vet before being transferred to the attending police officer when available.

Prior to wider adoption of new methods data needs to be reviewed, disseminated, and validated following community guidelines. To date, CDnaRP data has been distributed through a series of reports, scientific papers, conference presentations, professional working group meetings, and training events. Ultimately the

CDnaRP will host these outputs through an online platform with links to online training materials for police, veterinary practitioners, and livestock keepers.

Conclusion

The CDnaRP was established as a forum to develop, promote and apply best practice guidelines for the collection and analysis of Canine DNA from livestock and wildlife attacks. Results have shown the UK Canine DNA analysis pipeline can be further optimized with data being confirmed through the distribution of the LAW DOG kits. Data resulting from kit use in livestock attacks is pending and will be used to further assess different methods and use by end-user groups prior to making recommendations for wider adoption of sampling methods in the future.

Acknowledgments

The authors would like to thank North Wales Police, National Police Chiefs Council (NPCC) Livestock Attack Working Group, Chartered Society of Forensic Science Animal Related Crime Working Group, National Wildlife and Rural Crime Units, Forensic Capability Network, National Sheep Association, Science and Advice for Scottish Agriculture (SASA) Wildlife Forensic Laboratory, Cellmark Forensic Services, NFU Mutual, Farmers Union Wales (FUW), Synergy Farm Health, IVC Evidencia.

Conflict of interest statement

The work described has been funded by DEFRA (AW1414) and through LJMU Research and Innovation Services.

References

[1] NPCC. Livestock Worrying Police Working Group Final report. Available at https://news.npcc.police.uk/resources/livestock-worrying-police-working-group-feb-2018. Accessed on 04.10.2024.

[2] NFU Mutual. UK cost of livestock worrying rises by nearly 30%. Available at https://www.nfumutual.co.uk/media-centre/complacency-kills-as-uk-cost-of-livestock-worrying-rises-by-nearly-30/. Accessed on 04.10.2024.

[3] Welsh Government Press Release. Cabinet Secretary welcomes new course addressing dog attacks on livestock in Wales. Available at https://www.gov.wales/cabinet-secretary-welcomes-new-course-addressing-dog-attacks-livestock-wales#:~:text=RDOC%20aims%20to%20support%20dog,roll%20out%20of%20this%20course. Accessed on 04.10.2024.

[4] Cheshire Constabulary. Operation Recall: new initiative launched to tackle livestock worrying. Available at https://www.cheshire.police.uk/news/cheshire/news/articles/2023/10/operation-recall-new-initiative-launched-to-tackle-livestock-worrying/. Accessed on 04.10.2024.

[5] Animal Welfare (Kept Animals) Bill. Available at https://bills.parliament.uk/bills/2880. Accessed on 04.10.2024.

[6] PAW Forensic Working Group. Guidance document for the DNA swabbing of animal bite marks and sampling from suspected offending animals. Available at https://www.tracenetwork.org/wp-content/uploads/2018/05/Swabbing-of-bites-guidance-document-final-July-2017.pdf. Accessed on 04.10.2024.

[7] Verdon TJ, Mitchell RJ, van Oorschot RA. Evaluation of tapelifting as a collection method for touch DNA. Forensic science international: Genetics. 2014 Jan 1;8(1):179-86.

[8] DEFRA. Fallen stock and safe disposal of dead animals. Available at https://www.gov.uk/guidance/fallen-stock. Accessed on 04.10.2024.

[9] Dawnay N. Optimising the collection of canine DNA from worried livestock for forensic identity testing. DEFRA Report AW1414. Available on request.

[10] Dawnay et al 2024. Development of a qPCR assay for the quantification of canine autosomal DNA recovered from livestock attacks. Science and Justice, in press.

[11] Dayton et al 2009. Developmental validation of short tandem repeat reagent kit for forensic DNA profiling of canine biological material. Croatian Medical Journal 50(3):268-85.

[12] Berger et al 2014. Validation of two canine STR multiplex-assays following the ISFG recommendations for non-human DNA analysis. Forensic Science International: Genetics 8(1):90-100.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 217-223
DOI: https://dx.doi.org/10.15304/cc.2025.1869

NanoSSID: a bioinformatics pipeline for species identification in forensic genetics

Mengyao Zhao[1], Xinyue Zheng[1], Yiping Hou[1], Suyu Li[1], Linchuan Liao[2], Zheng Wang[1], Jing Liu[1,*]

[1] *Institute of Forensic Medicine, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu 610041, China.*

[2] *Department of Forensic Toxicological Analysis, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu 610041, China*

** Corresponding author at: No.24, South Section 1, Yihuan Road, Chengdu, Sichuan, China.*

Email: liujing2381@scu.edu.cn

Tel: +86-28-85501549; Fax: +86-28-85501550

Abstract

Species identification has special forensic value in many cases, such as combating wildlife smuggling and tracking food adulteration. The mitochondrial 12S rRNA gene has sequence information characteristics of low intra-species variation and high inter-species variation, making it suitable for species identification. In the present study, the 12S rRNA gene sequences of over 100 different species obtained from public database were used to develop an automated nanopore sequencing-based species identification pipeline, called NanoSSID. We sequenced 64 species and compared their sequencing results with each other to evaluate sequence similarity, and the NanoSSID was used to determine whether the detected species can be correctly assigned to the corresponding species. In addition, we assessed the ability of this pipeline to distinguish samples mixed with different species. Most species exhibited over 80% similarity, and some species showed 95% to 99.5% similarity with 1 to 20 base differences. Species with significant differences can be correctly assigned to their respective species by the NanoSSID, while species with high similarity can only be identified by examining the different bases among these species. Furthermore, this pipeline can successfully identify different components in mixed samples and showed the corresponding mixing ratio trend. Overall, the NanoSSID exhibited strong species identification capability and can play an effective role in the future.

Keywords

Forensic genetics; 12S rRNA gene; Species identification; Nanopore sequencing; *NanoSSID*

1. Introduction

Species identification plays an important role in many cases, such as wildlife smuggling, animal attacks, and food adulteration [1]. Conventional approaches based on morphology or immunology are easy to operate but incapable of precise detection on a large scale. With the rapid development of DNA technology, DNA barcoding is becoming a preferred method for species identification using the standardized regions of DNA sequences [2]. Some mitochondrial genes have been selected as DNA barcodes due to the characteristics of multicopies per cell, lack of recombination, and accelerated evolutionary rate [3]. The mitochondrial 12S rRNA gene has sequence information characteristics of low intra-species variation and high inter-species variation [4], making it suitable for species identification.

Currently, the gold standard for DNA barcoding is Sanger sequencing, but it is difficult to distinguish individual components when faced with mixed or contaminated samples [5]. Nanopore sequencing has the advantages of ultra-long read length, low cost, portability, and the ability to identify mixed samples [6]. Several bioinformatic pipelines for species identification based on Oxford Nanopore Technologies's MinION platform are currently available [7, 8], demonstrating the great potential of nanopore sequencing in this field. In addition to MinION, Qitan Technology launched its commercial nanopore sequencer QNome-3841 in 2021, which has been initially explored in several fields in recent years [9, 10]. Therefore, in this study, we developed a species identification bioinformatic pipeline called NanoSSID based on the *12S rRNA gene sequences* and explored the applicability of this pipeline using the QNome nanopore platform.

2. Material studied, methods, techniques

2.1 NanoSSID development

The 12S rRNA gene reference sequences of over 100 different species obtained from public database were used to develop an automated nanopore sequencing-based species identification pipeline, called NanoSSID (Fig. 1). FastANI was utilized to compare the sequence similarity between species.

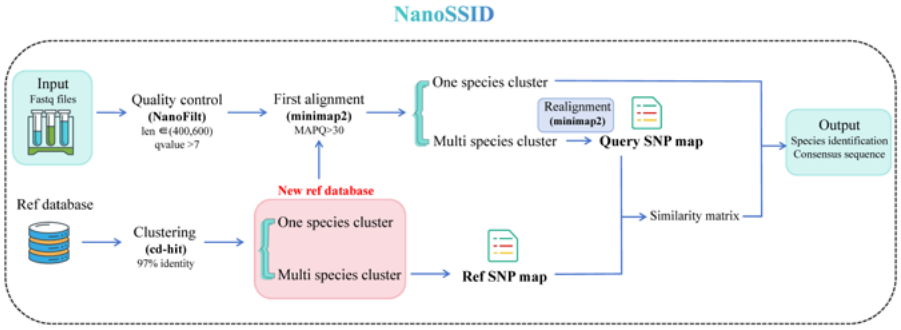


Fig. 1. Schematic diagram of a bioinformatics workflow for species identification using QNome sequencing data. *The corresponding biometric software for each step is shown in parentheses.*

2.2 Sample preparation

This study was approved by the Guangzhou Wildlife Research Center in China (ethical license: GZZOO20210903A) and conducted under the ethical principles of the "Helsinki Declaration" of the World Medical Association. We selected 64 species as test samples and purchased beef, chicken, duck, and pork from the market to prepare simulated mixed samples with vary proportions and ingredients. All samples underwent DNA extraction using the TIANamp Genomic DNA Kit (Tiangen Biotech) and subsequently conducted PCR amplification with 12S rRNA universal primers. The PCR products were confirmed by agarose gel electrophoresis and quantified using the Qubit 3.0 fluorometer (Thermo Fisher Scientific).

2.3 QNome nanopore sequencing and Sanger sequencing

The library was prepared using the QDL-E V1.1 kit (Qitan Technology), which involved the end-repair process, purification of end-repair products, ligation of sequencing adapters, purification of ligation products, and quantification. The normalized library was sequenced on the QNome-3841 sequencer (Qitan Technology) with the Qcell-384 sequencing chip (Qitan Technology) and the QDS V1.1 kit (Qitan Technology). The model "QDSv2.0_QCell384Pv2.0_QDLEv1.0_NTGv" was applied for base calling.

3. Results

3.1 Reference database performance

In general, most species exhibited over 80% similarity, and some species showed 95% to 99.5% similarity with 1 to 20 base differences. Based on a 97% sequence similarity threshold, reference sequences were clustered into 85 one-species clusters and 14 multi-species clusters. The sequences between one-species clusters differ widely, and can be directly distinguished based on the alignment results. However, within multi-species clusters, the sequence similarity between species is extremely high, specific recognition of the differential loci between species is required.

3.2 QNome nanopore sequencing performance

The raw sequencing data was filtered by Q-score and read lengths. After that, the average number of total reads was 563,993 (minimum: 154,071; maximum: 1,324,232). The average mapped reads was 562,992 (minimum: 154,049; maximum: 1,324,109), and the ratios of effective reads (mapped reads/total reads) were mostly above 99%, and the lowest was 97.25%. Overall, it presented an adequate amount of data and good sequencing quality.

3.3 NanoSSID identified test samples

Among the test samples, 88.33% of samples successfully matched to the reference sequences of the same name, whereas 11.67% of samples matched to the reference sequences of closely related species for various reasons. The average proportion of sequencing reads assigned to the reference sequence was 95.66%, with a maximum of 100%. In general, NanoSSID can almost correctly determine the origin of the species, even for those with extremely high sequence similarity.

3.4 NanoSSID identified mixed samples

We mix different proportions of chicken, duck, and pork into beef to simulate meat fraud. The average proportion sequencing reads assigned to the reference sequence was 98.87%, with an average sequence identity of 99.60%. All samples with different mixing ratios were successfully detected all expected species and could reflect the changing trend of mixing ratios (**Fig. 2**).

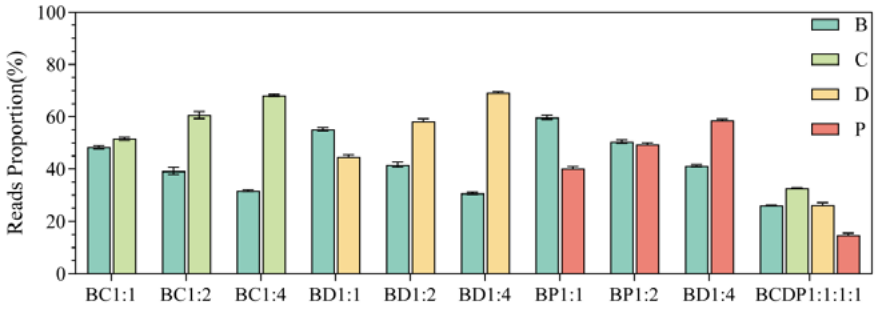


Fig. 2. Distribution diagram of read ratios for mixed samples. B represents beef; C represents chicken; D represents duck; P represents pork.

4. Discussion

Some of the test samples were not successfully matched by NanoSSID to the reference sequences of the same name, primarily due to several reasons. Some samples were unable to find a matching reference sequence in the public database and could only be compared to closely related species. Several samples demonstrated identities significantly below 98% with their corresponding reference sequences, suggesting possible mislabeling. Additionally, certain samples may be hybrid species or exhibit special intraspecific variations, showing high similarities with more than one reference sequence. Moreover, some samples exhibited mixed bases in their consensus sequences, which may be related to sequencing errors or mitochondrial DNA heterogeneity.

Although simulated mixed samples reflected changes in mixing ratios to a certain extent, there were significant differences between the read proportion and fresh weight basis of the samples, mainly due to the following biological and technical factors. First of all, under the same mass, different species and various tissues of the same species exhibit variability in both cell size and quantity, as well as the variable amount of mitochondria in the cells [11]. Secondly, manual measurement errors may occur during weighing and mixing. In addition, the loss during DNA extraction process and the efficiency of PCR amplification can also fluctuate. Therefore, the quantitative results obtained through NanoSSID analysis can only be regarded as a rough estimate of the relative content of different species. Overall, despite some limitations, NanoSSID meets the identification needs of most species and mixed samples with a convenient process.

5. Conclusion

In this study, based on the the 12S rRNA gene sequences of over 100 different species, we constructed NanoSSID, an automated nanopore sequencing-based species identification pipeline. This pipeline exhibited strong species identification capability, which can correctly assigned species with sequence similarity up to 99.5% to the corresponding species, and successfully identify different components in mixed samples. It can play an effective role for combating wildlife smuggling and tracking food adulteration in the future.

6. Acknowledgments

This study was supported by grants from the Major Program of National Natural Science Foundation of China (82293653) and the National Natural Science Foundation of China (81871532). The funders had no role in study design, data collection and analysis, preparation of the manuscript, or decision to submit the manuscript for publication.

7. Conflict of interest statement

The authors declare that they have no conflict of interest.

8. References

[1] M. Staats, A.J. Arulandhu, B. Gravendeel, et al., Advances in DNA metabarcoding for food and wildlife forensic species identification, Anal Bioanal Chem 408(17) (2016) 4615-4630.

[2] S. Dobrovolny, M. Blaschitz, T. Weinmaier, et al., Development of a DNA metabarcoding method for the identification of fifteen mammalian and six poultry species in food, Food Chemistry 272 (2019) 354-361.

[3] H. Nakamura, T. Muro, S. Imamura, et al., Forensic species identification based on size variation of mitochondrial DNA hypervariable regions, Int J Legal Med 123(2) (2009) 177-184.

[4] L. Yang, Z. Tan, D. Wang, et al., Species identification through mitochondrial rRNA genetic analysis, Sci Rep 4(1) (2014) 4089.

[5] R.-R. Xing, R.-R. Hu, J.-X. Han, et al., DNA barcoding and mini-barcoding in authenticating processed animal-derived food: A case study involving the Chinese market, Food Chemistry 309 (2020) 125653.

[6] D. Plesivkova, R. Richards, S. Harbison, A review of the potential of the MinION™ single-molecule sequencing system for forensic applications, WIREs Forensic Science 1(1) (2019) e1323.

[7] J. Kissenkötter, S. Böhlken-Fascher, A. Abd El Wahed, Flesh ID: Nanopore Sequencing Combined with Offline BLAST Search for the Identification of Meat Source, Foods 9(10) (2020) 1392.

[8] N. Vasiljevic, M. Lim, E. Humble, et al., Developmental validation of Oxford Nanopore Technology MinION sequence data and the NGSpeciesID bioinformatic pipeline for forensic genetic species identification, Forensic Science International: Genetics 53 (2021) 102493.

[9] K. Peng, Y. Yin, Y. Li, et al., QitanTech Nanopore Long-Read Sequencing Enables Rapid Resolution of Complete Genomes of Multi-Drug Resistant Pathogens, Front. Microbiol. 13 (2022).

[10] J. Liu, S. Li, Y. Su, et al., A proof-of-principle study: The potential application of MiniHap biomarkers in ancestry inference based on the QNome nanopore sequencing, Forensic Science International: Genetics 68 (2024) 102947.

[11] G. Cottenet, C. Blancpain, P.F. Chuah, et al., Evaluation and application of a next generation sequencing approach for meat species identification, Food Control 110 (2020) 107003.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 225-232
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Development and validation of a fluorescence PCR assay for the species identification of ten animals

Ruocheng Xia[1], Ruxin Zhu[1], Ruiyang Tao[1*], Kaijun Ma[2*]

[1] *Shanghai Key Laboratory of Forensic Medicine, Shanghai Forensic Service Platform, Academy of Forensic Sciences, Key Laboratory of Forensic Science, Ministry of Justice, Shanghai 200063, China.*

[2] *Shanghai Key Laboratory of Crime Scene Evidence, Shanghai Research Institute of Criminal Science and Technology, Shanghai 200072, China.*

** Corresponding author at: No. 1347 West Guangfu Road, Putuo District, Shanghai, taoruiyang163@163.com; makaijun@sina.com.*

Abstract

Objetives To develop a rapid and reliable multiplex fluorescence PCR assay for the simultaneous identification of ten different animal species (goat, cattle, donkey, chicken, pig, horse, sheep, goose, mouse, and duck). **Materials and methods** Based on the complete mitochondria genome sequences of the ten animals in the GenBank database, a comparative genomics analysis strategy was used to screen DNA sequences with both intra-species conservation and inter-species specificity. Species-specific primers were designed to construct a multiplex fluorescence PCR assay that can simultaneously identify ten animal species. Subsequently, the performance of the assay, including species specificity, sensitivity, reproducibility, and concordance, was validated for use in forensic analysis according to the guidelines of the Scientific Working Group on DNA Analysis Methods (SWGDAM). In addition, simulated mixed meat samples were also detected using this assay. **Results and conclusions** Ten pairs of species-specific primers were designed to amplify partial regions of the mitochondria genome (D-loop, 16S rRNA, NADH2, Cytb, COI, ATP6, NADH4, and NADH6) of the ten animals. The assay was constructed by combining these primers for species identification through multiplex fluorescence PCR and capillary electrophoresis. The results of developmental validation indicated that this new assay is reproducible, accurate and reliable. Specifically, the species specificity study showed no cross-reactivity with a variety of common animal species or humans. The sensitivity study demonstrated a detection limit of 31.25 pg for the ten species. Furthermore, the assay successfully detected duck components as low as 0.5% in mixtures of duck and beef. The developed multiplex fluorescence PCR assay effectively achieves species identification for ten common animals, making it a valuable tool for detecting meat fraud and adulteration.

Keywords

species identification, meat adulteration, mitochondrial DNA, multiplex PCR, capillary electrophoresis.

1. Introduction

Meat adulteration has emerged as a global challenge that poses a significant threat to food safety and consumer trust [1,2]. Given the gravity of the situation, there is an urgent need for the development of rapid and reliable detection methods that can accurately identify meat adulteration, ensuring the safety and authenticity of meat products and safeguarding consumer health and interests.

In recent years, species-specific multiplex PCR based on the amplification of mtDNA genes has received extensive attention, which involves the amplification of different fragments of mtDNA genes using multiple pairs of species-specific primers, followed by electrophoresis [3-6].

In the present study, we describe a simple and rapid method for the simultaneous identification of ten animal species (goat, cattle, donkey, chicken, pig, horse, sheep, goose, mouse, and duck) based on mtDNA genes, using multiplexed fluorescent PCR followed by capillary electrophoresis. Subsequently, the assay was validated for species specificity, sensitivity, repeatability, and consistency according to SWGDAM guidelines. The study aims to provide a reliable detection method and solution strategy for identifying adulteration in common meats and their products.

2. Materials and methods

2.1. Sample collections and DNA extraction

Maet samples from ten species (goat, cattle, donkey, chicken, pig, horse, sheep, goose, mouse, and duck) used for the construction of a multiplex PCR system were collected from local retailers and markets in Shanghai. Maet samples from nine species (human, cat, dog, rabbit, monkey, pigeon, snake, frog, and fish) used in the species-specific study were collected from local retailers and markets in Shanghai, except for the commercial DNA sample 9948 (Promega, Madison, WI, USA), which was used as the human reference sample.

DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Quantification of the DNA

was performed using the Qubit™ dsDNA HS Assay Kit with the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

2.2. Selection of species-specific sequences and primer design

Mitochondrial reference genome sequences for the ten target species were downloaded from the GenBank database (http://www.ncbi.nlm.nih.gov/genbank). DNA sequences exhibiting intra-species conservation and inter-species specificity were identified using MEGAX software (https://www.megasoftware.net/). Specific primers were designed with Primer Premier 5.0 software (Premier Biosoft International, Palo Alto, CA, USA).

2.3. Construction of the multiplex fluorescent PCR system

The primer sets were systematically grouped based on the fragment lengths of the amplification products and assigned to one of two fluorescent labeling dyes: 6-FAM (blue) or HEX (green) at the 5' end. All primers were synthesized and labeled by Sangon Biotech Co. Ltd. (Shanghai, China).

All fluorescent primers were combined, and the final concentration of each primer, along with the optimal annealing temperature and cycle number, was optimized based on the genotyping results. Notably, both positive and negative controls were included in each PCR reaction.

2.4. Capillary electrophoresis and genotyping

All samples were separated by capillary electrophoresis using an ABI 3500XL Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) equipped with POP-4 polymer and a 36 cm capillary array (Applied Biosystems, Foster City, CA, USA). Briefly, the loading samples for capillary electrophoresis consisted of 1 µL of PCR product, 8.5 µL of Hi-Di formamide (Applied Biosystems, Foster City, CA, USA), and 0.5 µL of the T-500 internal size standard. Genotyping data for all samples were analyzed using GeneMapper™ ID-X software (Applied Biosystems, Foster City, CA, USA), with a threshold of 50 RFU for peak analysis.

2.5. Species specificity study

DNA samples from nine non-target species (human, cat, dog, rabbit, monkey, pigeon, snake, frog, and fish) were tested using the new multiplex assay to determine the presence of any cross-reactivity.

2.6. Sensitivity study

A serial dilution of positive control DNA was amplified in triplicate with quantities of 2 ng, 1 ng, 500 pg, 250 pg, 125 pg, 62.5 pg, and 31.5 pg. The mean peak heights and percentage of detected alleles were calculated for each DNA template.

2.7. Mixture study

Ground duck was mixed with beef at mass ratios of 0.5%, 1%, 5%, and 10%, with the total mass of each mixed sample being 100 mg. DNA from each mixed sample was then extracted, detected, and analyzed in triplicate.

3. Results

3.1. Construction and Optimization of the multiplex PCR system

In this study, we ultimately selected gene sequences from ten species for designing species-specific primers, including the *Capra hircus* D-loop, *Bos taurus* 16S rRNA, *Equus asinus* NADH2, *Gallus gallus* Cytb, *Sus scrofa* COI, *Equus caballus* ATP6, *Ovis aries* NADH4, *Anser cygnoides* 16S rRNA, *Mus musculus* NADH4, and *Anas platyrhynchos* NADH6.

Based on the genotyping results of the positive control samples, the concentration of each primer in the multiplex assay was optimized. Subsequently, the PCR reaction conditions for the multiplex assay were refined. The total volume of the PCR system was 15 μL, consisting of 7.5 μL of 2× Multiplex PCR Master Mix (Qiagen, Hilden, Germany), 1.5 μL of 5× Q-Solution (Qiagen, Hilden, Germany), 1 μL of a forward and reverse primer mixture (10 μmol/L), 1 ng of extracted DNA, and deionized water to make up the remaining volume. The PCR reaction was performed using the GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA) under the following conditions: pre-denaturation at 95°C for 15 minutes, denaturation at 94°C for 30 seconds, annealing at 58°C for 90 seconds, and extension at 72°C for 90 seconds, for a total of 30 cycles, followed by a final extension at 60°C for 60 minutes. The final genotyping profile is shown in Figure 1.

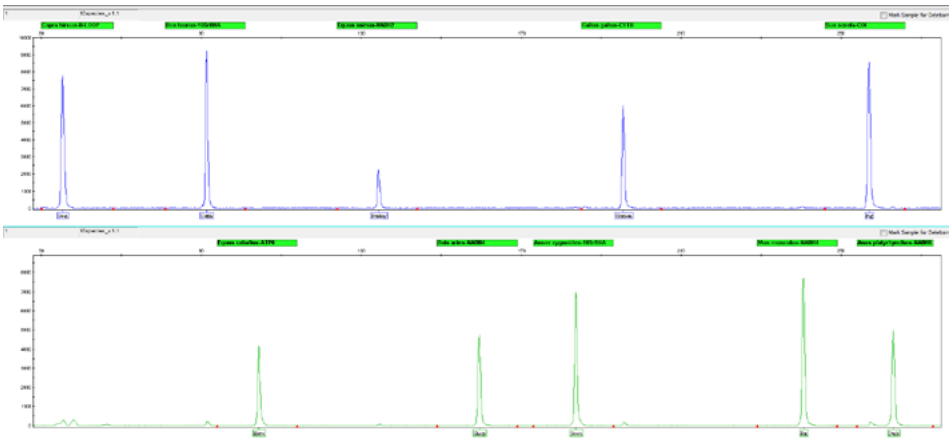


Figure 1. Electropherogram of positive control DNA (equal amounts of DNA were mixed from each species and then diluted to 1 ng/μL) with the multiplex.

3.2. Species specificity study

In the species specificity study, no reproducible peaks above 50 RFU were observed from the human, cat, dog, rabbit, monkey, pigeon, snake, frog, or fish DNA templates.

3.3. Sensitivity study

In the sensitivity study, complete genotyping profiles were obtained with positive control DNA inputs ranging from 2 ng to 31.25 pg, with average peak heights ranging from 6516 RFU to 452 RFU (Figure 2).

3.3 Mixture study

In the mixture study, we prepared four samples with beef and duck mass ratios of 0.5%, 1%, 5%, and 10%. The results showed that duck meat was detectable even when its mass was as low as 0.5 mg.

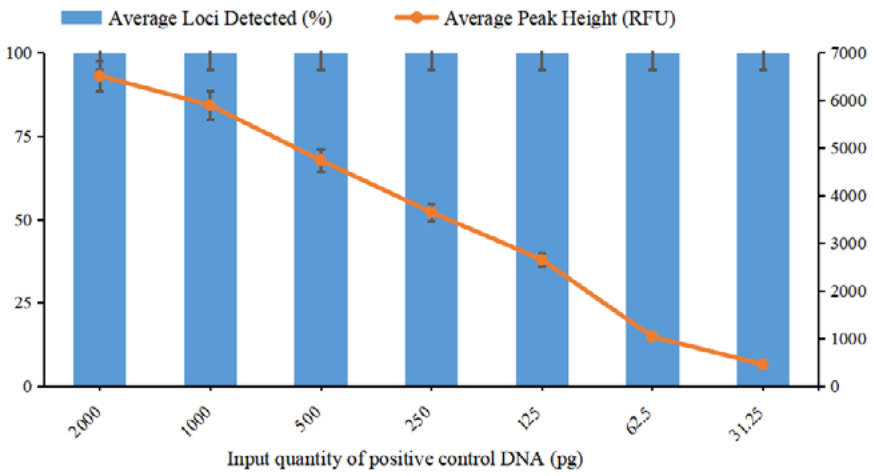


Figure 2. Sensitivity testing of template DNA ranging from 2 ng to 31.25 pg. The left Y-axis represents the average percentage of loci detected, and the right Y-axis represents the average peak height. Error bars show the SDs between three replicates.

4. Discussion

In this study, we successfully developed a simple and rapid method for animal species identification based on mitochondrial DNA, utilizing multiplex PCR combined with capillary electrophoresis. We specifically targeted nine of the most commonly consumed food species in China: goat, cow, donkey, chicken, pig, horse, sheep, goose, and duck. Additionally, rat meat, which, despite being inedible and associated with health risks, was included to further enhance the system's practical applicability.

The design of species-specific primers was based on key mtDNA genes, including D-loop, 16S rRNA, NADH2, Cytb, COI, ATP6, NADH4, and NADH6. After optimizing primer concentrations, annealing temperatures, and PCR cycling conditions, we established a multiplex fluorescence PCR assay capable of simultaneously identifying multiple species in a single reaction. This not only improves the efficiency of detection but also provides a streamlined solution for species identification in forensic and food safety contexts.

The assay was validated in accordance with the rigorous guidelines set by the SWGDAM. In the species specificity study, we confirmed that all primers

demonstrated high specificity for their respective species, with no cross-reactivity observed between closely related species or other non-target animals. Despite multiplex PCR generally being 10 to 100 times less sensitive than simplex PCR [7], the assay developed in this study was able to detect DNA quantities as low as 31.25 pg, a level sufficient for many practical applications.

Moreover, the results of the mixture study showed that the assay could reliably detect positive signals even in samples adulterated with as little as 0.5% duck meat, underscoring its robustness and sensitivity in real-world applications. This reliable detection of mixed samples highlights the system's utility for preventing meat fraud and ensuring the integrity of food products, where even minimal adulteration must be identified.

5. Conclusion

This study describes the development and validation of a new rapid and reliable multiplex fluorescent PCR assay for the simultaneous identification of ten different animal species: goat, cattle, donkey, chicken, pig, horse, sheep, goose, mouse, and duck. Following the guidelines published by SWGDAM, validation studies were conducted, demonstrating that the assay possesses high accuracy, sensitivity, and species specificity. Additionally, the new system indicated reliable detection of mixed meat samples, making it a valuable tool for detecting meat fraud and adulteration.

6. Acknowledgments

This study was supported by the Opening Project of Shanghai Key Laboratory of Crime Scene Evidence (2023XCWZK02), and the Public Interest Research Grant Programs of National Research Institutes (GY2024Z-2). The funding institutions had no role in study design, data analysis, publishing decisions, or manuscript preparation.

7. Conflict of interest statement

None.

8. References

[1] M. Kaltenbrunner, W. Mayer, K. Kerkhoff, et al. Differentiation between wild boar and domestic pig in food by targeting two gene loci by real-time PCR. Sci Rep. 9,1 (2019) 9221.

[2] J.C. Li, J.P. Li, S.G. Xu, et al. A rapid and reliable multiplex PCR assay for simultaneous detection of fourteen animal species in two tubes. Food Chem. 295 (2019) 395-402.

[3] M.E. Ali, M.A. Razzak, S.B.A. Hamid, et al. Multiplex PCR assay for the detection of five meat species forbidden in Islamic foods. Food Chem. 177 (2015) 214-24.

[4] M. Razzak, J. Keramat, M. Kadivar, et al. Multiplex-PCR As a Rapid and Sensitive Method for Identification of Meat Species in Halal-Meat Products. Recent Pat Food Nutr Agric. 8,3 (2017) 175-182.

[5] P.Z. Qin, W. Qu, J.G. Xu, et al. A sensitive multiplex PCR protocol for simultaneous detection of chicken, duck, and pork in beef samples. J Food Sci Technol. 56,3 (2019) 1266-1274.

[6] A. Galal-Khallaf. Multiplex PCR and 12S rRNA gene sequencing for detection of meat adulteration: A case study in the Egyptian markets. Gene. 764 (2021) 145062.

[7] N. Ishida, M. Sakurada, H. Kusunoki, et al. Development of a simultaneous identification method for 13 animal species using two multiplex real-time PCR assays and melting curve analysis. Leg Med (Tokyo). 30 (2018) 64-71.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 233-239
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Using the microbiome in soil to locate the origin of soil stains

Aurora Hofsvang[1,2,*], Eirik Natås Hanssen[1], Eric Jacques de Muinck[2], Ane Elida Fonneløp[1,2]

[1] *Department of Forensic Sciences, Oslo University Hospital, Oslo, Norway*
[2] *Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway*

** Corresponding author at: Oslo University Hospital- Rikshospitalet, 0424 Oslo, Norway.*
Email: auhofs@ous-hf.no

Understanding how soil microbiomes can be used to link objects or individuals to specific locations could facilitate the development of a new method in forensics. This exploratory study investigates how well machine learning can predict the origin of soil and soil stains by analyzing microbial composition. Soil was collected from 13 locations in and around the Oslo area, and a piece of clothing stained with soil was created for each place. The sampling was performed in a manner that included both spatial and temporal variation of the locations in the dataset. The soil samples were predicted with 92% accuracy while the soil stain samples were predicted with 90% accuracy.

Keywords

Microbiomics, Geolocalization, Soil forensics

Introduction

As the sensitivity of human DNA analysis has improved, there has been an increasing need for methods able to provide context for human DNA evidence in forensic science. One such method is using the microbiome in soil to link a suspect or an object to a specific location, so-called geolocalization. Soil is easily transferable and is in use as evidence in some countries today [1]. Microbiomics may be an improvement upon traditional soil forensic methods as studies suggest that microbial composition of soil varies both with geographic distance and ecological and

environmental differences [1]. This makes the method not only capable of distinguishing between soil types but also between similar soils from different places. In addition, the method relies on DNA extraction and sequencing analysis, which is becoming methodically established in forensics today, limiting the need for separate equipment and experts.

Several studies have investigated various aspects of this new method, such as transferability, temporal variation, and stability over time [2], but more research is needed. This study aims to establish if a soil stain with a limited amount of soil is still suited for analysis and if machine learning can be used to predict the origin of soil and soil stains based on microbial composition.

Methods

Soil samples were collected from 13 sites in and around Oslo, Norway. At each site sampling was performed 2-3 times on a weekly basis with two samples at each timepoint, resulting in a total of 85 samples. The soil samples were collected by placing a sampling tube straight into the ground thereby collecting the top few cm of soil. In addition, soil stains on fabric (mock clothing samples) were created by rubbing a piece of fabric against the ground at the exact sampling site. The participants were instructed to rub the fabric through their fingers before sampling, to mimic a worn piece of clothing. All the samples were placed in a -20 freezer for storage until further analysis, except three of the fabric samples (from locations C, F, and L) that were stored at room temperature for around 1.5 months.

DNA was extracted from approximately 250µl of the collected soil and from swab samples of the fabric soil stains, using the MagAttract® PowerSoil® DNA Kit (Qiagen). The library preparation was performed according to a protocol described by de Muinck et al [3]. It involved two PCR reactions to amplify the V4 region of the 16S gene and triple-index the samples. Pair-end sequencing with a read length of 300bp was performed on a MiSeq (Illumina) using V3 reagents.

The raw reads were demultiplexed and trimmed. DADA2 [4] was used to denoise the data to amplicon sequence variants (ASVs) and for taxonomy assignment using the Silva reference database [5]. All DADA2 work was done on a High-Performance Computing cluster (Sigma2). RStudio (version 2024.04.1+748) was used for further analysis, including multidimensional Scaling to investigate sample clustering (NMDS), regression analysis to examine the effect of different factors (PERMANOVA), and cross-validation to investigate prediction performance (Random forest).

Results

The focus of the analysis was to visualize and understand the variation and composition of the dataset and then investigate the origin location of the samples using machine learning.
After filtration, 17 059 ASVs distributed amongst 76 samples from 13 locations remained. A total of 9 samples were lost during various filtering steps in the method. Thus, there were between four to seven samples from each location.

Initial data exploration

Differences in the microbial composition between samples were calculated using Bray-Curtis dissimilarity. The distance matrix was then plotted using NMDS (Figure 1), showing that samples from different locations were distinct, though the degree of difference varied. For some locations, the soil stain sample appears relatively different from the soil samples from the same location (for example for locations C, K, and Q), but overall it seems like sample type (soil or soil stain on clothing) is not a significant driver of clustering.

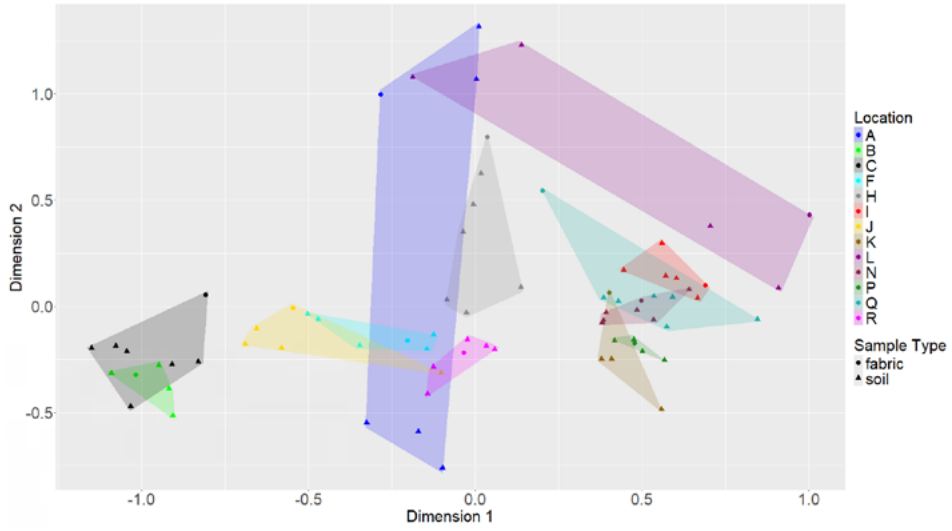


Figure 1: An NMDS plot to visualize the relative differences between microbial composition in samples as distance in two dimensions. The points in the plot represent different samples which are colored after location (A-C, F, H-L, N, P-R) and shaped based on sample type (soil or soil stain).

The effect of sampling location, sampling time, and sample type on the structure of the distance matrix was further investigated using Permutational Multivariate Analysis of Variance (PERMANOVA). The results revealed that location had a significant influence on the clustering of the samples across all pairwise comparisons (p-values <0.05), and sample type contributed significantly to differences between samples for the pairwise comparisons between locations A and H, A and K as well as H and K. However, these results must be interpreted cautiously, as there were significant differences in within-group dispersion between some locations and only one soil stain sample was collected from each location, reducing the reliability of these pairwise comparisons.

Prediction of origin location

The dataset was separated based on sample type (soil or soil stain on clothing). This was done to first train a model and test the prediction only on soil samples, and then compare this result to the accuracy of a model trained with soil samples predicting the origin location of soil stains. Machine learning was used to predict the source location of soil samples based on microbial composition. Random Forest was used with 5-fold cross-validation 5 times, each time with a different seed. This was done to ensure each sample was predicted with models based on different compositions of training data. In total 315 predictions were made (5*63 samples) and 292 of them were correct, giving a 92% prediction accuracy (figure 2a). Samples from nine of the 13 locations were predicted with 100% accuracy. When training on soil samples and predicting on fabric samples a slightly lower prediction accuracy of 90% was achieved (figure 2b). Out of the 13 soil samples on fabric, four were predicted wrongly and only one of these (fabric from location Q) was predicted wrong by all models.

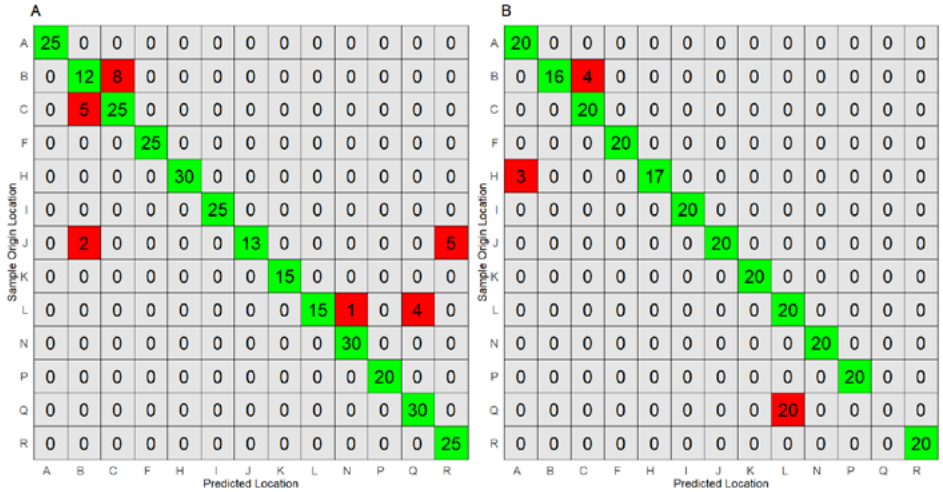


Figure 2: Confusion matrix from prediction of soil samples (A) and fabric samples (B). The y-axis shows the location of the predicted samples, and the x-axis shows the predicted location of the samples.

Discussion

In this study we explored differences in the microbial composition in soil from different locations and if we can exploit these differences to predict the origin of samples using machine learning.

Our results indicate that there are significant differences between the microbiome in soil at different locations, as the soil from the 13 sites we investigated showed distinction in the microbiome through PERMANOVA and with NMDS. This is supported by other research findings in that the microbiome in soil is very diverse and that the diversity is connected to location [1].

PERMANOVA and NMDS were also used to investigate if samples from soil stains significantly differ from samples from soil in microbial composition. Our results indicate that there are some differences in the microbiome, but that overall sample type has less impact on the separation of the samples than location. This aspect needs more investigation with a larger sample size to determine if observed differences stem from outliers or systematic differences because of sample type. Soil stain samples were collected on fabric on which participants deposited their skin microbiome. A future study to determine if the observed differences between

soil samples and soil stain samples stem from the additional skin microbiome would be useful. This could also explain some of the inconsistencies in the impact of sample type between locations, as participants may deposit different amounts of microorganisms on the fabric.

The prediction by Random Forest of origin location for soil and soil stains was overall very successful with high accuracy for both predictions of soil and soil stains. In a criminal case it is reasonable to assume that the evidence will not be a scoop of soil, but rather a soil stain on clothing. The high accuracy achieved when predicting the origin of soil samples in this study indicates that these types of samples can be suited for this analysis.

Although the high prediction accuracies are encouraging, some of the samples showed ambiguous prediction results, indicating that some model optimization is still possible and needed. Small changes in the shuffling of training data had an impact on which samples that were predicted wrong, even for the prediction on soil stain samples where the test data was constant. This can be caused by for example overlap in microbial composition from different places or noise in the data making it hard for the model to find the real underlying microbial compositional signatures. In microbial research high dimensionality is often a challenge and variable selection is suggested as part of the best practice [6]. Some studies suggest that this could help stabilize the model [7].

Conclusion

In this study we found that soil collected from 13 different locations in and around the Oslo area significantly differs in microbial community composition, and that machine learning could be used to predict the origin location of the soil based on these differences with a 92% accuracy for soil samples and with 90% accuracy for fabric samples. The high accuracy of the Random Forest predictions indicates that soil and soil stains could be suited for analysis of microbiomes to connect evidence to a specific location. However, further research is needed to fully understand the possibilities and limitations, including studying influencing factors such as participant differences and variation over time, as well as optimizing the prediction model, for example, through variant selection.

Acknowledgments

We would like to acknowledge the volunteers participating in the sampling of soil.

Conflict of interest

The authors declare no conflict of interest

References

[1] Habtom, H., et al., *Applying microbial biogeography in soil forensics.* Forensic Science International: Genetics, 2019. **38**: p. 195-203.

[2] Robinson, J.M., et al., *Forensic Applications of Microbiomics: A Review.* Frontiers in Microbiology, 2021. **11**.

[3] de Muinck, E.J., et al., *A novel ultra high-throughput 16S rRNA gene amplicon sequencing library preparation method for the Illumina HiSeq platform.* Microbiome, 2017. **5**(1): p. 68.

[4] Callahan, B.J., et al., *DADA2: High-resolution sample inference from Illumina amplicon data.* Nat Methods, 2016. **13**(7): p. 581-3.

[5] Pruesse, E., et al., *SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB.* Nucleic Acids Research, 2007. **35**(21): p. 7188-7196.

[6] Asnicar, F., et al., *Machine learning for microbiologists.* Nature Reviews Microbiology, 2024. **22**(4): p. 191-205.

[7] Marcos-Zambrano, L.J., et al., *Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment*. Frontiers in Microbiology, 2021. **12**.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 241-247
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Standardization of Molecular Analysis Techniques for DNA Identification of Bird Species from Eggshells

Caio Max Araujo[1,3], Gabriella Maia Nantes da Silva[1,2], Maria Alice S. Alves[3], Dayse Aparecida da Silva[1*]

[1]*Laboratório de Diagnósticos por DNA, Dept. Ecologia/IBRAG, Universidade do Estado do Rio de Janeiro, Rio de Janeiro/Brasil.*

[2]*Instituto Federal de Educação, Ciência e Tecnologia do Mato Grosso Do Sul, Campus Três Lagoas/Brasil.*

[3]*Laboratório de Ecologia de Aves e Comportamento, Dept. Ecologia/IBRAG, Universidade do Estado do Rio de Janeiro, Rio de Janeiro/Brasil.*

** Corresponding author at: Laboratório de Diagnósticos por DNA, Dept. Ecologia/IBRAG, Universidade do Estado do Rio de Janeiro, Rio de Janeiro/Brasil, dayse.a.silva@gmail.com,*

Abstract

Advancements in genetic analyses have been increasingly contributing in our understanding of ecological, evolutionary, and phylogenetic aspects of animals. Examples of this progress can be seen in the forensic field. In Brazil, wildlife traffic stands out as one of the most prevalent environmental crimes, with significant impacts on avian populations. For these animals, obtaining genetic material through blood is more advantageous due to the presence of nucleated red blood cells. However, this method can induce stress and alter the behavior of these animals, proving more challenging for threatened and elusive species. Consequently, non-invasive techniques for obtaining genetic material have been increasingly studied. The present study aimed to standardize molecular analysis techniques in bird eggshells. For standardization, incubated eggshells of Common Gallinule (Gallinula galeata) were used, from which fragments were collected weekly over 30 days post-hatch and separated into three categories ("shell,""membrane," and "shell + membrane"). DNA was extracted and purified using sodium acetate, and the quantity and quality of the molecule were evaluated by fluorimetry and 0.8% agarose gel electrophoresis. Polymerase chain reaction (PCR) amplifications of the nuclear marker Chromodomain Helicase DNA Binding Protein (CHD) and the mitochondrial marker Cytochrome B (CytB) were conducted for genetic determinations. Forty-five samples were analyzed, from which DNA extraction was successful in 91% (n=41). DNA from shells was not detected after the third week post-hatch. The highest DNA concentration was obtained in membranes (2.074±2.073 µg/ml). Success rates for markers were 51.1% for CytB (n=23) and 44.4% for CHD (n=20). No amplification was observed for CytB after

three weeks post-hatch (n=9), regardless of the biological material used. The tested methodological conditions were suitable for molecular analysis from eggshells with and without membranes, with DNA extraction possible even four weeks post-hatch, albeit with a lower success rate for CytB amplifications. We hope to conclude this observation for a larger number of samples and markers.

Keywords

DNA extraction, Birds identification, Conservation, Forensics

1. Introduction

Illegal wildlife trade ranks among the five most profitable illicit activities in the world [1], and high demand makes the capture and trade of wild animals a significant factor in the decline of populations and ecological services of species subjected to this pressure [2]. Birds have been used as pets, decoration, and even in traditional medicine [3], representing one of the most coveted groups in the illegal trade, with some families being more exploited, such as Thraupidae and Psittacidae. In Brazil, birds are the most trafficked group among other taxa, with species like the saffron finch (*Sicallis flaveola*) and the double-collared seedeater (*Sporophila caerulescens*) being the most trafficked [4].

The use of genetic tools has generated new knowledge about animal ecology and evolution, providing important information on population genetics [5], evolutionary history [6], and even insights into extinct species [7]. The most efficient way to obtain genetic material from birds is through their blood, as they have nucleated red blood cells [8]. However, obtaining this material can potentially cause stress and alter the animal's behavior [9], which could have even greater consequences for cryptic or endangered species.

Due to this, researchers have been seeking non-invasive methods of obtaining genetic material [10, 11]. Materials such as feathers, eggshells, and even feces have been used for various purposes, such as species identification [12], conservation [13], and forensic studies [14]. The present study aimed to standardize molecular analysis techniques in bird eggs over time.

2. Material studied, methods, techniques

2.1. Gallinula galeata *eggshell's*

Five eggs of Gallinula galeata were collected at Rodrigo de Freitas Lagoon and later incubated at the UERJ Bird Ecology and Behavior Laboratory until the

chicks hatched. Only three chicks hatched from the eggs and were subsequently returned to their respective nests.

Eggshells were stored in open boxes, and fragments were collected weekly, starting one day after hatching and continuing for four weeks post-hatching (T0-T4). The fragments were categorized into three groups: fragments of shells without membrane, fragments of shells with membrane [15], and fragments of membrane (hereafter referred to as shell, shell + membrane, and membrane, respectively). Subsequently, the fragments were ground (using a porcelain mortar and pestle) and stored in a freezer until further analysis.

2.2. DNA extraction and amplification

DNA extraction was performed using ammonium acetate [16]. To the powdered eggshell, 500 µL of lysis solution (0.5 mL Tris-Cl pH 7.5 1M, 4 mL EDTA pH 8.0 0.5M, 2 mL NaCl 5M, 40 mL SDS 10%) was added, along with 15 µL of proteinase K (0.5 mg/mL) and 15 µL of DTT (30 mM) for digestion at 55ºC, 800 rpm for 16 hours. The supernatant was extracted and combined with 193 µL of ammonium acetate (2M). 500 µL of isopropanol (100%) and 52.5 µL of sodium chloride (0.25M) were used for DNA isolation and precipitation. The concentration of the extracted DNA was quantified by spectrophotometry using a Qubit. Samples were amplified for the markers Cytochrome B (CytB), responsible for species identification, and Chromodomain Helicase DNA Binding Protein (CHD), which allows for identifying the animal's sex.

3. Results

DNA was successfully extracted from 91% of the samples (n=41). Among the samples from which DNA could not be extracted, two were from the third week and two from the fourth week after hatching. The samples had an average of 44 mg of shell, 45 mg of shell + membrane, and 4 mg of membrane. The average concentrations for the samples were 0.023µg/mL, 0.618µg/mL, and 2.074µg/mL for shell, shell + membrane, and membrane, respectively.

For the CytB marker, amplifications were successful up to the third week post-hatching, with a total success rate of 51.1% (n=23), including two from shell samples, eleven from shell + membrane, and ten from membrane. The selected sequences for sequencing were correctly identified as *Gallinula galeata* when compared to sequences in GenBank.

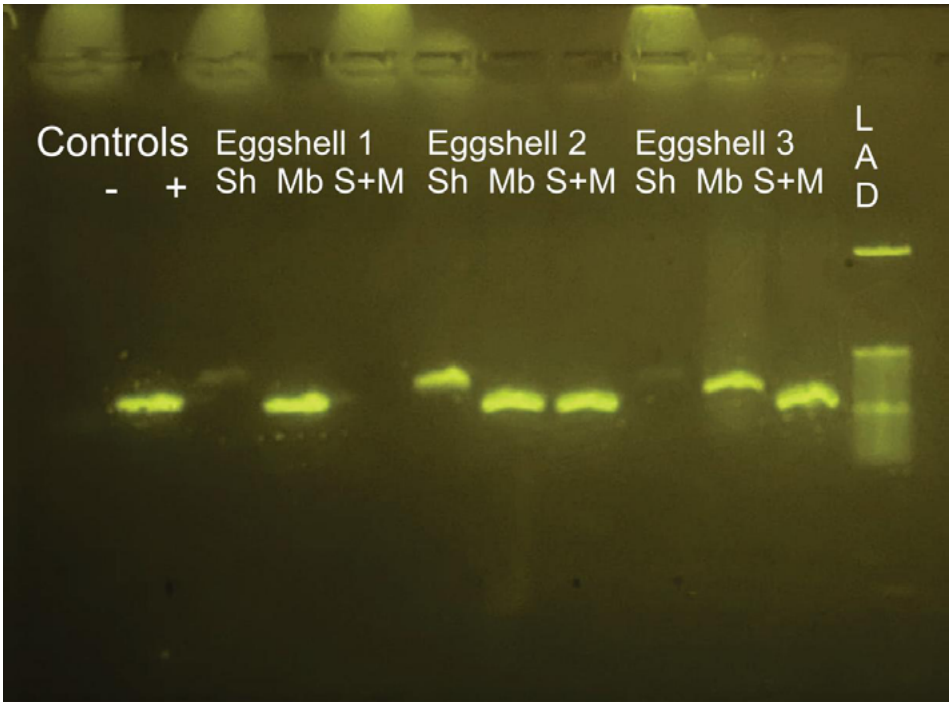


Figure 1: CytB amplifications, one day after eclosion (T0). Controls negative (-), positive (+), shell (Sh), membrane (Mb), shell + membrane (S+M) and ladder (LAD).For the CHD marker, although samples were amplified at all tested time points, it showed a success rate of 44.4%, amplifying only 20 samples, including one from shell, seven from shell + membrane, and twelve from membrane. Only one sample was identified as female, while all others were male, showing the pattern of a single amplification band.

4. Discussion

Despite having the lowest weight among the samples, the eggshell membrane proved to be the best material for obtaining DNA, with a concentration nearly 100 times greater than the shell and three times greater than the shell + membrane. This result may be associated with remnants of blood vessels from the extraembryonic membranes adhered to the eggshell membrane [15]. Interestingly, the mixture of shell + membrane showed similar results to samples containing only the membrane regarding the amplification efficiency for CytB and CHD. The low amplification success for the shell may be linked to the scarcity of cells, as this structure is rich in mineral substances [17]. Nevertheless, obtaining DNA from the

shell is possible, as parental cells from the brood, through blood remnants, can adhere to the outer surface of the eggs after laying [9]. Another possibility for the presence of DNA in this structure is the deposition of dead epithelial cells from the parents during incubation [9].

The result obtained from the CHD marker indicated that the extracted DNA belonged to the chick, as all results identified males. The only sample that yielded a female result came from the membrane. In this case, we believe the sample may have been contaminated during the separation of the material from the shell, as this would be the only possible source for the female material.

5. Conclusion

The results show how these used biological samples can be useful for obtaining genetic material from birds. The best samples would be those containing some part of the membrane either adhered to or isolated from the egg shell. Although they have a lower success rate, shell fragments can also be used for DNA extraction. For greater success in species identification, we recommend using the material up to two weeks after hatching, as the amplification success rate for CytB is higher during this period. The methodological conditions tested were suitable for molecular analysis of eggshells with and without membranes, and we hope to conduct further observations for a more significant number of species and markers.

6. Acknowledgments

We would like to acknowledge Isabela Moraes, Rafael Saint'Clair and the members of the Laboratory of Bird Ecology and Behavior at UERJ for their logistical support with the collection, incubation, and return of the chicks to the Rodrigo de Freitas Lagoon. Alexandru Stermin helped with the use of the incubator. We also thank Igor Josué, Daniela Almeida, Laís Oliveira, João Marreiro, Patrícia Domingues, and the entire team at the DNA Diagnostics Laboratory at UERJ for their instruction and assistance with molecular techniques. Additionally, we extend our gratitude to the funding institutions, the Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro (FAPERJ) and the National Council for Scientific and Technological Development (CNPq) for the financial support provided to M.A.S.A (CNE #201126/2022 and PQ#308.615/2022-0, respectively) and to D.A

(PQ# 306208/2020-2). C.M.A has been granted by an undergraduate fellowship from UERJ (PIBIC/UERJ).

7. Conflict of interest statement

None.

8. References

[1] S.M. Barber-Meyer, Dealing with the clandestine nature of wildlife-trade market surveys, *Conserv. Biol.* 24 (2010) 918-923.

[2] C. Dai, C. Zhang, The local bird trade and its conservation impacts in the city of Guiyang, Southwest China, *Reg. Environ. Change* 17 (2017) 1763-1773.

[3] R.R.N. Alves, R.C.L. Leite, W.M.S. Souto, D.M. Bezerra, A. Loures-Ribeiro, Ethno-ornithology and conservation of wild birds in the semi-arid Caatinga of northeastern Brazil, *J. Ethnobiol. Ethnomed.* 9 (2013) 14.

[4] J.M. Ferreira, N. Moraes-Barros (Eds.), O tráfico de fauna silvestre segundo as notícias: avaliação das informações publicadas para o Brasil, Argentina e Paraguai, 1st ed., Freeland Brasil, São Paulo, 2023.

[5] G. Thom, A.T. Xue, A.O. Sawakuchi, C.C. Ribas, M.J. Hickerson, A. Aleixo, C. Miyaki, Quaternary climate changes as speciation drivers in the Amazon floodplains, *Sci. Adv.* 6 (2020)

[6] R. Batista, U. Olsson, T. Andermann, A. Aleixo, C.C. Ribas, A. Antonelli, Phylogenomics and biogeography of the world's thrushes (Aves, Turdus): new evidence for a more parsimonious evolutionary history, *Proc. R. Soc. B* 287 (2020) 20192400.

[7] C.L. Oskam, J. Haile, E. McLay, P. Rigby, M.E. Allentoft, M.E. Olsen, et al., Fossil avian eggshell preserves ancient DNA, *Proc. R. Soc. B* 277 (2010) 1991-2000.

[8] D. Martín-Gálvez, J.M. Peralta-Sánchez, D.A. Dawson, A.M. Martín-Platero, M.Á. Martínez-Bueno, T. Burke, J.J. Soler, DNA sampling from eggshell swabbing is widely applicable in wild bird populations as demonstrated in 23 species, *Mol. Ecol. Resour.* 11 (2011) 481-493.

[9] G. Schmaltz, C.M. Somers, P. Sharma, J.S. Quinn, Non-destructive sampling of maternal DNA from the external shell of bird eggs, *Conserv. Genet.* 7 (2006) 543-549.

[10] L. Begovic, I. Mihic, T. Pospihalj, T. Mikuska, S. Mlinaric, A. Mikuska, Evaluation of methods for molecular sex-typing of three heron species from different DNA sources, *Turk. J. Zool.* 41 (2017) 593-598.

[11] F. Seyedasgari, B. Asadi-Moghaddam, S. Sebestyen, R. Guillen, Feasibility of using non-invasive samples for early sex determination in falcon chicks by polymerase chain reaction, *Int. J. Conserv. Sci.* 13 (2022) 991-1002.

[12] G. Chilton, M.D. Sorenson, Genetic identification of eggs purportedly from the extinct Labrador Duck (Camptorhynchus labradorius), *Auk* 124 (2007) 962-968.

[13] T.A. Maia, S.T. Vilaça, L.R.D. Silva, F.R. Santos, G.P.D.M. Dantas, DNA sampling from eggshells and microsatellite genotyping in rare tropical birds: Case study on Brazilian Merganser, *Genet. Mol. Biol.* 40 (2017) 808-812.

[14] N.E. White, R. Dawson, M.L. Coghlan, S.R. Tridico, P.R. Mawson, J. Haile, M. Bunce, Application of STR markers in wildlife forensic casework involving Australian black-cockatoos (Calyptorhynchus spp.), *Forensic Sci. Int. Genet.* 6 (2012) 664-670.

[15] P. Yuda, A.W. Saputra, Eggshell membrane for DNA sexing of the endangered Maleo (*Macrocephalon maleo*), *F1000Res.* 9 (2020).

[16] D.A. Silva, V.M. de Oliveira, D. de Almeida, J. de S. Pinto, P.A. Succo, G.H.F. Seixas, M.A.S. Alves, Standardization of molecular analysis techniques for DNA identification of bird species of the Psittacidae family, *Forensic Sci. Int. Genet. Suppl. Ser.* 8 (2022) 105-107.

[17] O. Karlsson, C. Lilja, Eggshell structure, mode of development and growth rate in birds, *Zoology* 111 (2008) 494-502.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 249-256
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Forensic genetics in the conservation of Guiana Dolphins (Artiodactyla: Cetacea: Odontoceti)

M. Laeta[1, 2, §, *], M. Weksler[2], D.A. Silva[1]

[1] *Laboratório de Diagnóstico por DNA, Programa de Pós-graduação em Biociências, Universidade do Estado do Rio de Janeiro - UERJ, Rio de Janeiro – RJ, Brazil*

[2] *Setor de Mastozoologia, Departamento de Vertebrados, Museu Nacional/Universidade Federal do Rio de Janeiro, Rio de Janeiro – RJ, Brazil*

[§] *Current Addresses: Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo – SP, Brazil & Department of Ecoscience, Aarhus University, Roskilde, Denmark*

**Corresponding author: mairalaeta@gmail.com*

Abstract

The Guiana dolphin, *Sotalia guianensis* (van Bénéden, 1864) is the most common cetacean found in strandings along the Brazilian coast and, as a result, is well-represented in zoological collections. This species' broad distribution, coupled with increasing anthropogenic activities, subjects it to various pressures such as habitat degradation, pollution, and accidental/incidental captures. Carcasses provide crucial biological data, including sex determination, but the rapid decomposition in tropical environments and the absence of pronounced skeletal sexual dimorphism complicate this task. Behavioral differences between females and males, juveniles and adults, as well as among distinct populations, further suggest that exposure to anthropogenic threats may vary due to differential habitat usage. In this study, we aim to develop a forensic protocol for DNA extraction from skeletal remains by evaluating the efficacy of various bone structures—specifically, the humerus, vertebra, sternum, and periotic-tympanic complex. DNA was extracted from these tissues using a phenol-chloroform protocol, and quantification was performed using the Qubit® fluorometer. Results indicated that the vertebra and humerus yielded the highest DNA concentrations, followed by the periotic-tympanic complex and sternum. Polymerase chain reaction (PCR) amplification of nuclear and mitochondrial DNA fragments was conducted based on established protocols for cetaceans and humans. Among our results, we highlight the amplification of mitochondrial DNA for cytochrome b for the vertebra, sternum and periotic-tympanic complex. The analysis of bone fragments is crucial, as these tissues are commonly available in zoological collections. The molecular data generated from this study

contributes to the systematics, taxonomy, and evolutionary understanding of *Sotalia guianensis*, with important conservation implications. Notably, this threatened species, faces significant challenges across its range, which includes eight of the twelve recognized Management Units along the Brazilian coast, as well as throughout Central and South America.

Keyword

DNA extraction, bone, marine mammals, conservation

Introduction

Among the cetaceans that inhabit Brazil's extensive coastline, the small odontocete, Guiana dolphin *Sotalia guianensis* (van Bénéden, 1864), is notable for its broad distribution in estuarine and coastal waters. Its range extends from the state of Amapá to Santa Catarina, including the Amazon estuary [9, 12]. This species typically reaches a maximum body length of 230 cm, a weight of up to 150 kg, and has a lifespan of approximately 33 years [2]. Its use of coastal habitats, however, exposes it to various anthropogenic threats, such as habitat degradation, pollution, and by catch in fishing activities [1]. According to the International Union for Conservation of Nature (IUCN) Red List of Threatened Species, the Guiana dolphin is classified as "Near Threatened" across its entire distribution, which encompasses twelve recognized Management Units (MUs), eight of which are located along the Brazilian coast [2, 4]. This distribution and the delineation of MUs align with the genetic diversity patterns documented in populations along the Brazilian coast [6].

The Guiana dolphin is the most frequently represented odontocete in Brazilian zoological collections, largely due to its high mortality rate and the relative ease of recovering stranded carcasses. The examination of skeletal remains from these specimens has significantly advanced understanding in various fields, such as habitat use, age estimation, taxonomy, evolutionary history, conservation strategies, and health assessments [2, 3, 6, 8]. Genetic studies have revealed higher haplotype and nucleotide diversity in individuals from the northern Brazilian coast, particularly in the Amazon estuary, underscoring the genetic complexity within this region [6].

The overlap between the Guiana dolphin's distribution and anthropogenic activities places significant pressures on the species, including habitat loss, pollution, and both accidental and incidental captures [1, 2]. The recovery of carcasses is critical for documenting certain biological characteristics, such as sex; however, rapid decomposition in tropical climates and the lack of pronounced skeletal sexual

dimorphism complicate these determinations. In this study, we aim to establish a forensic protocol for DNA extraction from the bone tissue of the Guiana dolphin. We evaluate the effectiveness of DNA extraction from various bone structures—namely, the humerus, vertebra, sternum, and periotic-tympanic complex—followed by molecular analysis using nuclear and mitochondrial DNA markers on a specimen of known sex.

Material and methods

Material

As part of this project, bone tissue samples were collected from the vertebra, sternum, humerus, and periotic-tympanic complex of a male Guiana dolphin (*Sotalia guianensis*), specimen MN83455, housed at the Museu Nacional, Universidade Federal do Rio de Janeiro (MN/UFRJ). These tissues were selected for analysis to evaluate their suitability for DNA extraction.

Methods

Forensic DNA Extraction

DNA was extracted from the dry-preserved bone fragments - vertebra, sternum, humerus, and periotic-tympanic complex - of the specimen. These fragments were pulverized at low temperatures, and approximately 40mg of bone powder from each structure was digested in a solution containing EDTA and proteinase K. The DNA extraction followed the Phenol-Chloroform method as described by Sambrook et al. (1989). The extracted and purified DNA samples were quantified using a Qubit® 2.0 Fluorometer (Life Technologies). To assess the quality and quantity of the DNA, the samples were subjected to electrophoresis on 2% genomic agarose gels, composed of Tris base, Acetic acid, and EDTA (TAE). Electrophoresis was performed at 83V for 60 minutes, and the DNA bands were visualized under an ultraviolet (UV) transilluminator.

Amplification of nuclear X and Y fragments

Nuclear DNA markers from the ZFX, ZFY, and SRY regions were tested for sex determination. Gene amplification was carried out using the Polymerase Chain Reaction (PCR), following established protocols for odontocetes. The primers used were ZFXY582, ZFX923, and ZFY767, producing fragments of approximately 364 and 230 base pairs, respectively [7]; SRY593 and SRY764, yielding fragments of

approximately 170 base pairs; and ZFY1204 and ZFY0097, amplifying fragments of about 300 base pairs [10]. To ensure the accuracy of sex determination, two additional *Sotalia guianensis* specimens - one female (MN 83438) and one male (MN 83450) - were used as positive controls.

PCR reactions were conducted in a 20 µL reaction mixture containing 2.0 µL of DNA, 0.5 µL of each primer (10 µM, forward and reverse), 5 µL of 10× buffer, 0.2 µL of Platinum™ Taq DNA Polymerase (Invitrogen™), and 12 µL of DNA-grade water. The cycling profile and annealing temperatures were based on the protocol described by Cunha [5], using a Veriti® 96-Well Thermal Cycler (Applied Biosystems). The cycling conditions were as follows: an initial denaturation step at 94°C for 3 minutes, followed by 38 cycles of denaturation at 92°C for 1 minute, annealing at 50°C for 1 minute, and extension at 72°C for 1 minute. A final extension was performed at 72°C for 5 minutes. To verify amplification, 4.0 µL of the PCR product (amplicon) was mixed with 2 µL of Safe Dye (Thermo Fisher®) and run on a 2% TAE agarose gel at 83V for 30 minutes. A Low Mass DNA Ladder (50 bp, Promega®, Madison, USA) was used as the molecular weight marker.

Amplification of mitochondrial cytochrome b *locus*

The mitochondrial DNA gene cytochrome b (CytB) was amplified via Polymerase Chain Reaction (PCR). The primer combinations used were as follows: Forward primer L14816 and Reverse primer H15173, producing a fragment of approximately 300 base pairs [13]. PCR reactions were conducted in a 15 µL reaction volume containing 2.0–5.0 µL of DNA, 1.0 µL of each primer (10 µM, forward and reverse), 5.0 µL of MasterMix (Invitrogen™), and DNA-grade water to make up the final volume. The cycling profile and annealing temperatures followed the protocol of Parson *et al*. (2000) using a Veriti® 96-Well Thermal Cycler (Applied Biosystems). The thermal cycling conditions were as follows: initial denaturation at 95°C for 11 minutes, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 50°C for 45 seconds, and extension at 72°C for 45 seconds, with a final extension at 72°C for 10 minutes. To verify amplification, an aliquot of 4.0 µL of the PCR product (amplicon) was mixed with 2 µL of Safe Dye (Thermo Fisher®) and run on a 2% TAE agarose gel at 83V for 30–45 minutes. A Low Mass DNA Ladder (1 kb, Promega®, Madison, USA) was used as the molecular weight marker.

Results

Forensic DNA Extraction

Extraction using the phenol-chloroform protocol provided the following concentrations, as measured by the Qubit® fluorometer: the vertebra yielded the highest DNA concentration at 1.78 µg/ml, followed by the humerus at 1.10 µg/ml. The sternum and periotic-tympanic complex exhibited lower DNA concentrations, with 0.43 µg/ml and 0.37 µg/ml, respectively (mean of three samples for each structure). These results demonstrate that the removal of a small bone fragment can effectively facilitate the extraction from various skeletal structures of the Guiana dolphin.

The extraction of these bone fragments did not compromise the morphology of most of the skeletal elements of this specimen, including the vertebra (~15 x 18 cm), sternum (~21.5 x 12 cm), and humerus (~7 x 4 cm), as small bone fragments were excised from well-defined regions. In contrast, the removal of fragments from the periotic-tympanic complex was challenging due to the structural hardness resulting from its mineralization, leading to the disjunction of this small bone structure (~3.5 x 2 cm).

Sex determination using nuclear markers

The effective primer pair identified for amplifying approximately 364 base pairs in females was ZFXY582 and ZFX923. However, the target specimen (MN 83455) did not yield amplification of its nuclear DNA for sex determination, nor did the male positive control (MN 83450). In contrast, the female positive control (MN 83438) demonstrated successful amplification of the target fragment. Notably, no other primer pairs produced successful results for the target specimen or the positive controls.

Amplification of mitochondrial cytochrome b *locus*

Amplification of mitochondrial DNA for cytochrome b identification of fragments from the bone structures of the Guiana dolphin was possible for vertebra, sternum and periotic-tympanic complex (Figure 1), through 'Polymerase Chain Reactions' (PCR) based on the literature for several species, including mammals [13].

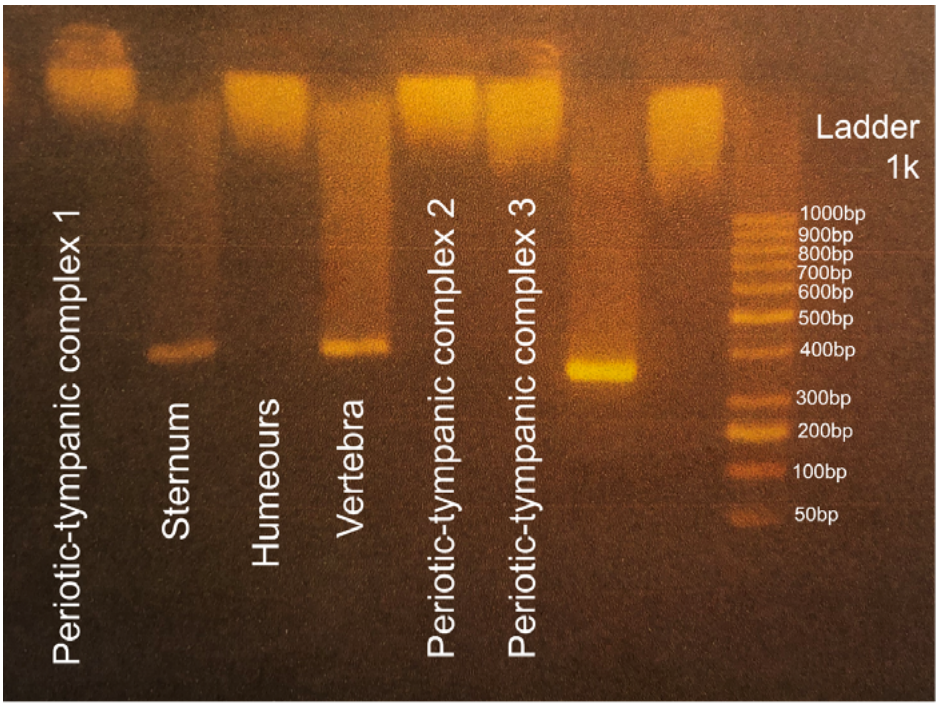


Figure 1. Example of gel with 'Polymerase Chain Reactions' (PCR) amplification products showing the effectiveness mitochondrial DNA extraction for some bone parts.

Discussion

DNA extraction from various parts of the skeleton of a specimen—initially found stranded on the beach and subsequently stored in a zoological collection—suggests that satisfactory preservation of genetic material has occurred, despite the lack of controlled processing conditions for this skeleton. Quantification of the DNA extracted using the phenol-chloroform method and measured with a Qubit® fluorimeter indicated that the less mineralized bone structures, specifically the humerus and vertebra, exhibited higher concentrations of DNA, followed by the sternum and periotic-tympanic complex. Among our results, we highlight the amplification of mitochondrial DNA for cytochrome b for the vertebra, sternum and periotic-tympanic complex using 'Polymerase Chain Reactions' (PCR).

The successful use of bone fragments for DNA extraction is particularly relevant for the estuarine-coastal Guiana dolphin, given that skeletons of this species

are widely available in zoological collections. Expanding this study, based on the forensic protocol established herein, may enhance our understanding of sex and population diversity among the Management Units of this dolphin [2, 6].

Differentiated management strategies for this taxon are crucial, as its distribution largely overlaps with human activities leading to environmental degradation and fishing practices, resulting in significant occurrences of accidental captures and bycatch [1, 2]. In Brazil, the Guiana dolphin is both the most abundant and the most threatened by anthropogenic impacts [4, 12]. The genetic structure of this species has revealed greater molecular diversity in the Amazon estuary compared to other regions along its extensive estuarine-coastal distribution [6].

Despite advancements in understanding this odontocete, further studies integrating genetic research and conservation efforts are necessary to inform effective management actions in the Brazilian Management Units [6]. The forensic protocol outlined in this study can facilitate effective DNA extraction from skeletal remains, contributing to the conservation management of *Sotalia guianensis*. These remains serve as vital records of biodiversity and environmental health, highlighting the importance of this sentinel species within estuarine-coastal ecosystems, particularly due to its role in bioaccumulation as a top predator in the food chain.

Conclusion

In conclusion, studies focused on this 'Near Threatened' estuarine-coastal dolphin emphasize the importance of utilizing data obtained from recovered carcasses to inform management and conservation strategies. By employing forensic techniques on the numerous skeletonized specimens available in zoological collections, researchers can enhance our understanding of the species' biology and ecology, ultimately contributing to more effective conservation actions for the Guiana dolphin (*Sotalia guianensis*).

Acknowledgments

Laboratório de Ciência e Tecnologia Forense-UERJ; Laboratório de Diagnósticos por DNA-UERJ; FAPESP 2023/11843-0 for the fellowship provided to M.L. FAPERJ and CNPq for the financial support provided to D.A.

Conflict of interest statement

The authors declare that they have no conflict of interest.

References

[1] A.P.M. Di Beneditto *et al.* Plano de ação nacional para a conservação do pequeno cetáceo Toninha: *Pontoporia blainvillei*. *Instituto Chico Mendes de Conservação da Biodiversidad* (2010).

[2] C. Domit *et al*. Progress report of the *Sotalia guianensis*. *In*: *International Whaling Commission Annual Scientific Committee Meeting* (2021).

[3] E. Monteiro-Filho *et al*. Skull shape and size divergence in dolphins of the genus *Sotalia*: a tridimensional morphometric analysis. *J. Mamm*. 83(2002): 125-134.

[4] E. Secchi, M.C.O. Santos, R. Reeves. *Sotalia guianensis*. *The IUCN Red List of Threatened Species* e.T181359A144232542(2018).

[5] H.A. Cunha. *Sistemática molecular e filogeografia do gênero Sotalia Gray 1866 (Delphinidae) no Brasil.* Doutorado. Universidade Federal do Rio de Janeiro (2007).

[6] H.A. Cunha *et al*. Review of population genetic data available to *Sotalia guianensis*. *In:* International Whaling Commission Annual Meeting (2020).

[7] M. Bérubé & P. Palsbøll. Identification of sex in cetaceans by multiplexing with three ZFX and ZFY specific primers. Mol. Ecol. 5(1996): 283-287.

[8] M. Laeta, S. Souza, S. Siciliano. Cervical ribs and cleft arches in brazilian *Sotalia guianensis*: contribution to comparative studies of skeletal morphology in norther and southeastern specimens. *Lat. Am. J. Aquat. Mamm.* 8(2010): 151-165.

[9] P.C. Simões-Lopes. Sobre a ampliação da distribuição do gênero *Sotalia* Gray, 1886 (Cetacea:Delphinidae) para as águas do Estado de Santa Catarina – Brasil. *Anais II Reunião de Trab. Esp. Mam. Aquát. Am. Sul*(1987).

[10] P.J. Palsbøll, A. Vader, I. Bakke, M. Raafat El-Gewely. Determination of gender in cetaceans by the polymerase chain reaction. *Can. J. Zool.* 70(1992): 2166-2170.

[11] S. Caballero *et al*. Taxonomic status of the genus *Sotalia*: species level ranking for "tucuxi" (*Sotalia fluviatilis*) and "costero" (*Sotalia guianensis*) dolphins. Mar. Mamm. Sci. 23(2007): 358-386.

[12] V.M. Da Silva & R. Best *Sotalia fluviatilis*. Mamm. Species. 396(1995): 1-7.

[13] W. Parson *et al*., 2000. Species identification by means of the cytochrome b gene. *Int. J. Leg. Med.* 114(2000): 23-28.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 257-264
DOI: https://dx.doi.org/10.15304/cc.2025.1869

The genetic male component of Afro-descendants from Pará (Brazil)

Julyana Ribeiro[1¥*], Lais Oliveira[1¥], Masinda Nguidi[1], Silvia Loiola[1], Milene Raiol[2], Ândrea Ribeiro-dos-Santos[2], Leonor Gusmão[1]

[1] *Laboratório de Diagnósticos por DNA (LDD), Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, Brazil.*

[2] *Laboratório de Genética Humana e Médica (LGHM), Universidade Federal do Pará (UFPA) Belém, Pará, Brazil.*

[¥] *These authors contributed equally to the work*

** Corresponding author at: Universidade do Estado do Rio de Janeiro (UERJ), Laboratório de Diagnóstico por DNA (LDD), Rua São Francisco Xavier, 524. Maracanã, 20550-900 Rio de Janeiro (RJ) Brazil, Tel.: +55(21)23342183, E-mail address: julyribeiro18.jr@gmail.com.*

Abstract

The present work aimed to increase our knowledge on the origin and diversity of paternal lineages of Quilombos from Pará (a State in the Amazonian region of Brazil). Therefore, 23 Y-STR markers were characterized in 226 individuals using the PowerPlex® Y23 system. A high haplotype diversity was observed. A comparative analysis between South American populations revealed greater proximity between admixed populations from Brazil, Colombia, and Afro-descendants from Brazil and Ecuador, all very close to Iberian populations. To infer continental contributions, haplogroups were predicted based on haplotypes. The high frequency of European paternal lineages (>67%) detected in this study was also observed in other studies of admixed populations from South America. However, contrary to expectations, given the history of the formation of Quilombos and their supposed isolation, paternal lineages of Indigenous origin were more frequent than those from Africa (~21% and >11%, respectively).

Keywords

Y-STR, Afro-descendant, Brazil, Haplogroup predictor.

Introduction

The genetic heterogeneity of current South American populations is the result of historical admixture events between Indigenous people, Europeans and Africans.

Brazil, in particular, was the country that maintained the highest proportion of Afro-diasporic population in the world, having received around 4.9 million Africans during the slavery period [1-2]. Although most of the current Brazilian population is admixed, there are native and Afro-Brazilian communities known as indigenous and Quilombos, respectively. Quilombos are communities formed by enslaved Africans and their descendants who resisted exploitation during slavery. Currently, Brazil has more than 2,807 communities legally certified as Quilombos, each with a particular history of formation. In recent years, studies using diverse markers revealed the high genetic heterogeneity of the Brazilian population, with different admixture patterns across country's regions. Genetic studies in Brazilian Afro-descendant populations can provide further information about the admixture processes behind the formation of current populations, and new insights into the genetic legacy of the African diaspora [3-6].

The present work aimed to contribute to the knowledge about the male lineage composition of South American populations using Y-chromosomal specific STR (Y-STRs). Thus, it was intended to provide relevant information to a database of Y-STR haplotypes to be used in subsequent population studies, as well as in forensic analyzes. Y-chromosomal specific markers are widely used in population genetics to infer paternal ancestry, as the non-recombining Y (NRY) portion is transmitted unchanged over generations in haplotype form, except in cases where mutations occur.

Material and methods

Bloodstain samples from 226 unrelated Afro-descendant males from Pará were collected in Whatman FTA® cards, under written informed consent. The project was approved by the ethical committee of the Health Sciences Institute of the Federal University of Pará (CAAE: 55906016.6.0000.0018). DNA extraction was performed using the Chelex resin [7].

The samples were genotyped using the PowerPlex Y23 kit (PPY23), following the manufacturer's instructions (Promega).

To predict the most likely haplogroup of each sample, the PPY23 haplotypes were submitted into NEVGEN Y DNA Haplogroup predictor (http://www.nevgen.org/). Haplogroup frequencies were calculated by direct counting.

The haplotype diversity (HD) was computed for the total population sample. Pairwise comparisons (F_{ST}) for 19 Y-STRs were performed between the

Afro-descendant (Pará), South American, and reference populations from the three main continents that contributed to South American admixture [8-15]. The Arlequin software ver. 3.5.1.2 [16] was used for HD and F_{ST} analyses. Pairwise genetic distances were visualized by Multi-Dimensional Scaling (MDS), as implemented in the STATISTICA software ver. 8.0 (www.statsoft.com).

Results

A high Y-STR haplotype diversity (HD = 0.9952 ± 0.0010) was found in the sample of 226 Afro-descendants from Pará, with 153 unique haplotypes identified. A total of 9 intermediate alleles were observed at the following *loci*: DYS448 (n = 1), DYS549 (n = 1), DYS570 (n = 5), DYS390 (n = 1), DYS458 (n = 13), and DYS385 (n = 1). These intermediate alleles were previously detected in other populations included in the YHRD.

Pairwise F_{ST} values *were calculated between Pará and other South American, African and European populations, and represented in a MDS plot (Fig. 1). The results showed a greater proximity between the Afro-descendants from Pará and other admixed populations from Brazil, Argentina, Colombia, and Paraguay* ($F_{ST} \leq 0.013$), *all genetically close to the Iberian populations* ($F_{ST} = 0.014$ *and* $F_{ST} = 0.023$, *Portugal and Spain, respectively). Significant differences were found when comparing Pará with the remaining populations* ($p < 5E\text{-}6$), *showing the largest* F_{ST} *values with the Africans* ($0.130 \leq F_{ST} \leq 0.172$) *and Indigenous* ($0.052 \leq F_{ST} \leq 0.170$), *followed by the two Afro-descendant populations from Esmeraldas (Ecuador) and Palenque (Colombia)* ($F_{ST} = 0.031$ *and* $F_{ST} = 0.058$, *respectively).*

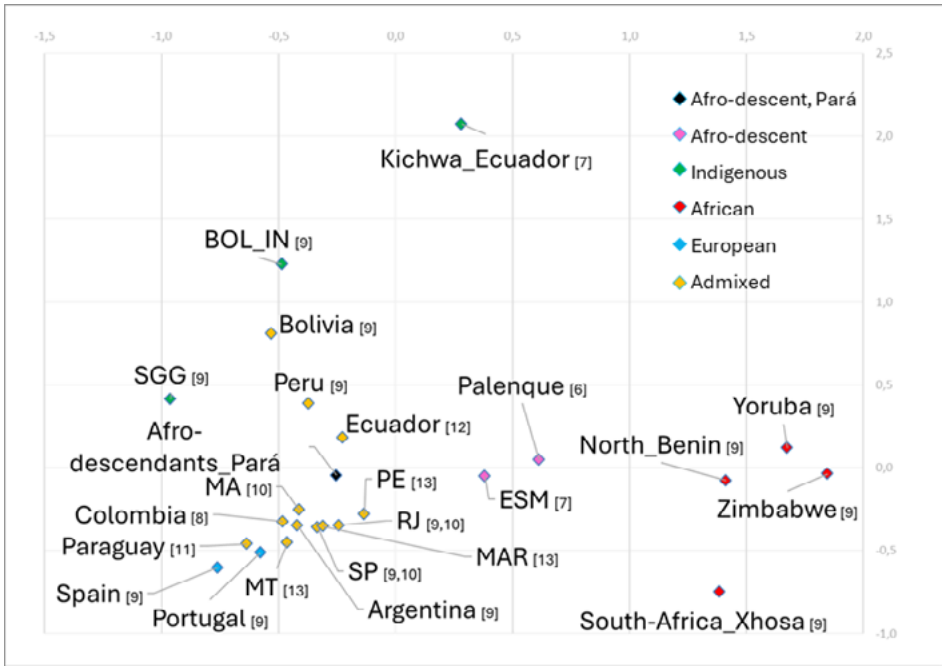


Fig. 1. *MDS plot of pairwise F_{ST} between Afro-descendant, admixed and Indigenous populations from South America, as well as African and European populations.*

Sample codes: ESM=Esmeraldas, Ecuador; BOL-IN=Indigenous people of Bolivia, Bolivia; MA=Manaus, Brazil; SGC=São Gabriel da Cachoeira, Brazil; PE=Pernambuco, Brazil; MAR=Maranhão, Brazil; MT=Mato Grosso, Brazil; RJ=Rio de Janeiro, Brazil; SP= São Paulo, Brazil. Stress = 0.0749950.

Haplogroups inferred from Y-STR profiles showed a high frequency of European lineages (65.93%), followed by Indigenous people (20.8%) and African lineages (11.5%) (Fig. 2). A single chromosome was found to belong to the Asian haplogroup O (0.44%). The most frequent haplogroup was R1b, representing 41.2% of the Afro-descendants from Pará.

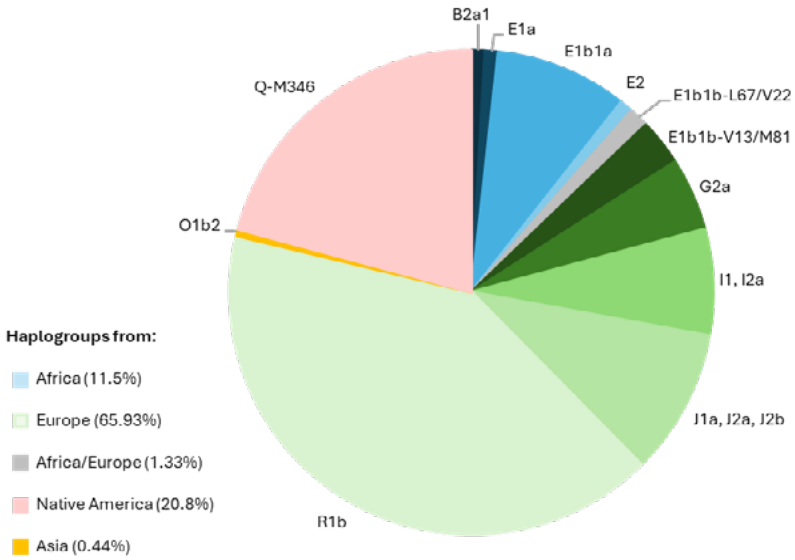


Fig. 2. *Haplogroup distribution inferred by NEVGEN in the Quilombos of Pará.*

Discussion

The high Y-STR haplotype diversity as observed in the Afro-descendants from Pará was earlier reported for Afro-descendant populations from Brazil, Colombia and Ecuador [5, 8-9].

Contrary to the expectations, given the history of the formation of the Quilombos and their supposed isolation, paternal lineages of Indigenous origin were more frequent than those from Africa. Haplogroups E1b1b-L67 and E1b1b-V22 lineages could be of either African or European origin. Therefore, the final European and African contributions were proportionally adjusted to 67.1% and 11.7%, respectively. In contrast with our findings, previous studies showed a prevalence of European and African lineages in most Brazilian Quilombos [17].

The R1b haplogroup is the most frequent lineage in Eastern Europe, reaching frequencies greater than 50% in the Iberian Peninsula [18]. It has also been reported at high frequencies in admixed South American populations [9, 15, 19]. These results are in line with expectations, as the predominance of male European

paternal lineages found in the Quilombos from Pará reflects the colonization history of South America by Europeans, mainly males.

Due to the error rate associated with haplogroup determination based on haplotype profiles, the results presented are just an approximation of the actual frequency of haplogroups in the studied population. For a more accurate definition of the haplogroups and their geographical origin, the results should be confirmed through Y-SNP genotyping.

Conclusion

With this study, we contributed to the forensic database of 23 Y-STR haplotypes from African-descent communities (Quilombos) from Pará, Brazil. The data generated provided valuable information about the paternal lineages of current populations in South America, allowing for a deeper understanding of the genetic composition and regional differences across the country. Contrary to expectations, the Quilombos of Pará presented a high European male contribution, followed by Indigenous people and African contributions. The results of the population comparison revealed a closer proximity between the Pará and other South American admixed populations, all very close to the Iberian populations, in agreement with the country's colonization history.

Acknowledgements

J. Ribeiro is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil (ref. 150075/2024-3 and ref. 32/2023). M. Nguidi is supported by CAPES, Brazil. Â Ribeiro-dos-Santos is supported by CNPq/Produtividade (ref. 304413/2015–1). L. Gusmão is supported by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Brazil (processes E-26/202.275/2019 and CNE-2022) and CNPq (ref. 315007/2023-0).

Conflict of interest statement

None

References

[1] P. Curtin, The Atlantic slave trade: a census, first ed., Wisconsin University Press, Madison, 1969.

[2] C. Moura, História do Negro Brasileiro, second ed., Ática, São Paulo, 1992.

[3] K. Abe-Sandes, W.A. Jr. Silva, M.A. Zago, Heterogeneity of the Y chromosome in Afro-Brazilian populations, Hum Biol. 76 (2004) 77-86.

[4] G.G. Ribeiro, R.R. de Lima, C.E. Wiesel et al., Afro-derived Brazilian populations: male genetic constitution estimated by Y-chromosomes STRs and AluYAP element polymorphisms, Am J Hum Biol. 21 (2009) 354-356.

[5] T.D; Palha, E.M. Ribeiro-Rodrigues, A. Ribeiro-dos-Santos et al., Male ancestry structure and interethnic admixture in African-descent communities from the Amazon as revealed by Y-chromosome Strs, Am J Phys Anthropol. 144 (2011) 471-478.

[6] I.A. Joerin-Luque, N.M. Sukow, I. D. Bucco et al., Ancestry, diversity, and genetics of health-related traits in African-derived communities (quilombos) from Brazil, Funct Integr Genomics 23 (1) (2023) 74.

[7] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a Medium for Simple Extraction of DNA for PCR-Based Typing from Forensic Material, Bio Techniques 10 (1991) 506-513.

[8] B. Martinez, F. Simão, V. Gomes et al., Searching for the roots of the first free African American community, Sci Rep. 10 (1) (2020) 20634.

[9] P. Villaescusa, M. Seidel, M. Nothnagel et al., A Y-chromosomal survey of Ecuador's multi-ethnic population reveals new insights into the tri-partite population structure and supports an early Holocene age of the rare Native American founder lineage C3-MPB373, Forensic Sci. Int. Genet. 51 (2021) 102427.

[10] S. Antão-Sousa, N. Pinto, P. Rende et al., The sequence of the repetitive motif influences the frequency of multistep mutations in Short Tandem Repeats, Sci Rep. 13 (1) (2023) 10251.

[11] J. Purps, S. Siegert, S. Willuweit et al., A global analysis of Y-chromosomal haplotype diversity for 23 STR loci, Forensic Sci. Int. Genet. 12 (2014) 12-23.

[12] I.B. Ambrosio, D.F. Braganholi, L.B.M. Orlando et al., Mutational data and population profiling of 23 Y-STRs in three Brazilian Populations, Forensic Sci. Int Genet. 48 (2020) 102348.

[13] J. Ribeiro, M. Romero, F. Simão, et al., Analysis of 23 Y-STRs in a population sample from eastern Paraguay, Forensic Sci. Int. Genet. 37 (2018) e20-e22.

[14] U. Toscanini, A. Gaviria, J. Pardo-Seco et al., The geographic mosaic of Ecuadorian Y-chromosome ancestry, Forensic Sci. Int. Genet. 33 (2018) 59-65.

[15] J. Jannuzzi, J. Ribeiro, C. Alho et al., Male lineages in Brazilian populations and performance of haplogroup prediction tools, Forensic Sci. Int. Genet. 44 (2020) 10216.

[16] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (3) (2010) 564-567.

[17] F.D.S. Gomes, Mocambos e Quilombos: Uma história do campesinato negro no Brasil, fisrt ed., Claro Enigma, São Paulo, 2015.

[18] K.L. Young, G. Sun, R. Deca et al., Paternal genetic history of the Basque population of Spain, Human Biology 83 (2011), 455-475.

[19] R. Resque, L. Gusmão, M. Geppert et al., Male Lineages in Brazil: Intercontinental Admixture and Stratification of the European Background, PLoS One 11(4) (2016) e0152573.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 265-273
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Qatar and the Middle East: Insights from a Y-STR Study Using PowerPlex Y23

Eida Khalaf Almohammed[1], Abdullah Hadi [2],Tarek Mghirbi [1], Khalid Marzook[1], Ibrahim Elramahi[1], Hayder Lazim[3*]

[1] *Ministry of Interior of Qatar, Forensic Medicine, Doha, Qatar*

[2] *Royal Preston Hospital, Sharoe Green Lane, Fulwood, Preston, PR2 9HT, UK*

[3] *Faculty of Health, Social Care and Medicine (FHSCM), School of Medicine, Edge Hill University, L39 4QP, UK*

** Corresponding author: Faculty of Health, Social Care and Medicine (FHSCM), School of Medicine, Edge Hill University, L39 4QP, UK (0044 7404988132) alazawihayder@yahoo.com, lazimh@edgehill.ac.uk (HL)*

Abstract

A large database on the Qatari population was created for comparison with other databases from the Arabian Peninsula and the Middle East. We provide a study of 23 Y-STR loci included in PowerPlex Y23 (Promega, USA) that were genotyped to produce haplotypes in 379 unrelated males from Qatar, a country at the crossroads of migration patterns. The graph of the population Q-matrix was created using Y-STR data from 134 Middle Eastern and African populations (11,305 individuals), and it revealed a stronger sub-grouping of populations within each ethnic group. The estimated migration rate between the Qatari and other Arabian populations was inferred using Bayesian coalescence theory in the Migrate-n program. According to the Gene Flow study, the main migration route was from Yemen to Kuwait through Qatar. Our research, using the PowerPlex Y23 database, shows the importance of gene diversity, as well as regional and social structuring, in determining the utility of demographic and forensic databases.

Keywords

Qatar, Middle East, Y-STR, Population genetics, Phylogenetics

Introduction

Qatar occupies the Qatar Peninsula, which extends northward from the Arabian Peninsula's eastern edge into the Arabian Gulf and covers 11,427 km2 (Fig 1).

Doha, the capital city, is located on the peninsula's eastern edge. Qatar has a population of around 3,005,069 million people, including 300,000 nationals of various origins [1]. Qatar has a southern border with Saudi Arabia and a maritime border with Bahrain, the United Arab Emirates, and Iran. It is known, like other countries in the region, for its distinct population structure, which is characterized by a high consanguinity rate [2, 3].

This study aims to investigate the Y-chromosomal genetic diversity of Qatar's Arab population. We analyze 23 Y-chromosomal short tandem repeat (Y-STR) loci using the PowerPlex Y23 system (Promega Corporation, Madison, WI, USA). Our objectives are: to elucidate the genetic structure of the Qatari population, to assess the genetic relationship between Qatari and other Middle Eastern populations and finally to shed light on historical human migration patterns across the Arabian Peninsula

This research contributes to our understanding of population genetics in the Middle East and may provide insights into the region's demographic history

1. Materials and methods

1.1. Qatari and the Middle Eastern populations samples

A total of 379 saliva samples were collected anonymously from unrelated indigenous Qatari males Arabs from Doha. Ethical approval was granted by the Ministry of Interior of Qatar and the University of Central Lancashire "Ref STEM454".

1.2. PCR amplification and fragment analysis

The PowerPlex Y23 kit was used to generate 23 loci. DNA amplification Reaction setup and thermal cycling were performed according to the procedures described in the PowerPlex Y23 kit User's Manual. Thermal cycling was performed in Veriti® 96-Well Thermal Cycler as recommended by the manufacturer (Thermo Fisher Scientific). Fragments were electrophoresed in eight capillaries (50-cm length) arrays on the ABI 3500 Genetic Analyser using the manufacturer's recommended protocols (Thermo Fisher Scientific) filled with POP-4™ polymer. GeneMapper IDX software V1.4 was used for allele calling and interpretation.

1.3. Statistical analyses

Molecular data were obtained for the Qatari population using Y-STRs based on the PowerPlex Y23 System and compared with the available data on other Middle Eastern populations [4]. Genetic distances between groups of males were quantified by the multi-dimensional scaling (MDS) plots which was generated by using R statistical software version 4.0. The phylogenetic tree was constructed using POPTREE2 online tool [5] and FigTree software [6].

The genetic structure of the Middle Eastern and different regions of Africa was investigated using the programme STRUCTURE version 2.3.7 and an admixture model [7, 8]. The Q-matrix graph of the studied populations was created using Y-STR data from 17 markers across 135 populations (11,305 individuals).

This study investigated gene flow between Qatari and other Middle Eastern populations. The Migrate software program employs a Bayesian inference framework to estimate several key population parameters, including effective population sizes, historical migration rates between varied populations under an asymmetrical migration model, and population divergences or admixture [9].

2. Results

The 23 Y-STRs profiled with PowerPlex Y23 kit were amplified for 379 Qatari males. Data are accessible from YHRD, release 69 (Accession number YA000494).

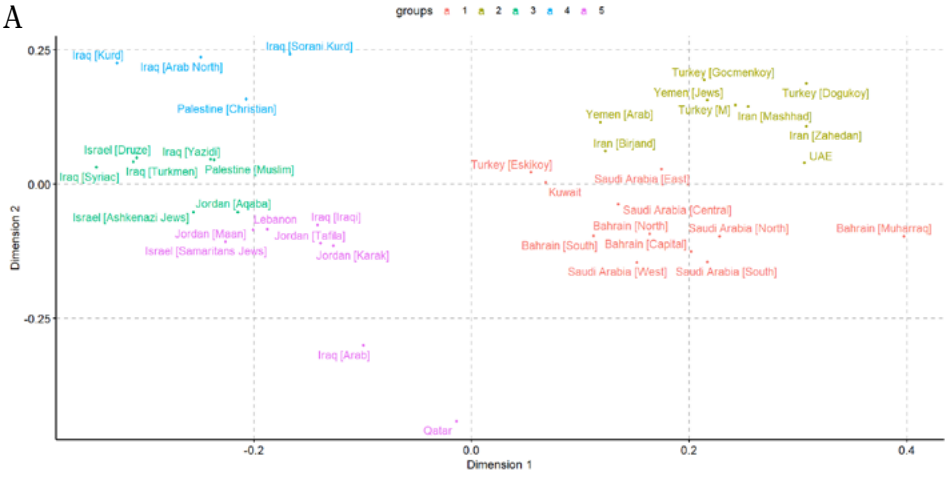
The MDS analysis was computed for all populations using the Y-filer 17-marker panel. The resulting MDS plot revealed five distinct clusters within the Middle Eastern population. Notably, the Qatari population did not cluster with other Arabian Gulf countries as might be expected. Instead, it exhibited greater genetic affinity with populations from the northern part of the Arabian Peninsula, the Levant region, and Iraq (Fig 1A).

Phylogenetic analysis identified four distinct genetic clusters (K=4) within the study population. As illustrated in Figure 1B, The Qatari population shared genetic characteristics with the most populated cluster in the region, which included 16 separate populations.

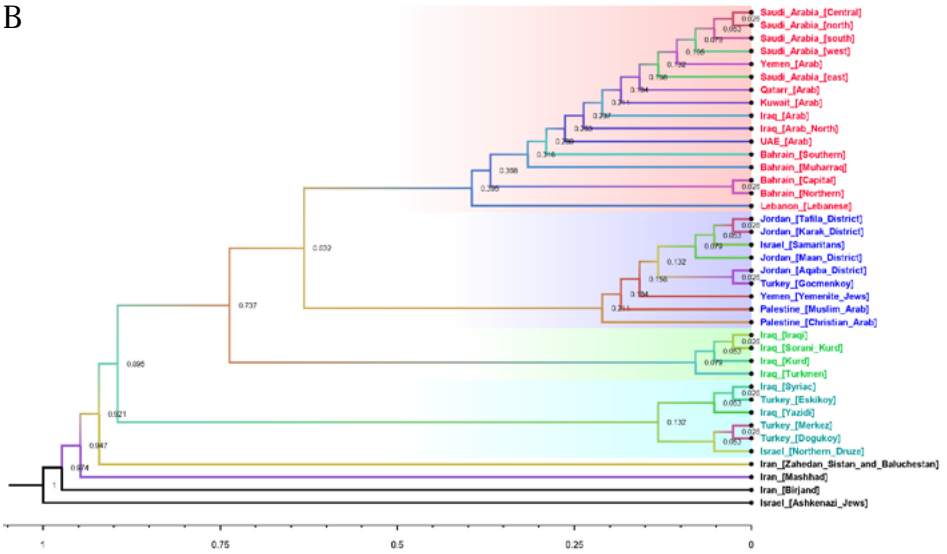
The Q-matrix graph (Fig 1C) showed eight clusters. Individual clusters in the Y-STR belonged to specific geographical regions without overlap, revealing a stronger sub-grouping of countries within each region. Clusters 2 and 4 were the most varied among Middle Eastern populations. While clusters 1 and 6 were the most varied among the African populations.

Migration rate analyses revealed two predominant migration routes within the Arabian Peninsula. The first route demonstrated significant gene flow originating from Qatar and extending to multiple destinations. Specifically, the highest migration rates were observed from Qatar to Yemen (Log(ml) = -2244.24), Saudi Arabia (Log(ml) = -2465.67), Kuwait (Log(ml) = -2384.68), and Iraq (Log(ml) = -2496.75). The second prominent migration route identified was originating from the UAE and terminating in Qatar (Log(ml) = -2444.35) (Fig 1D).

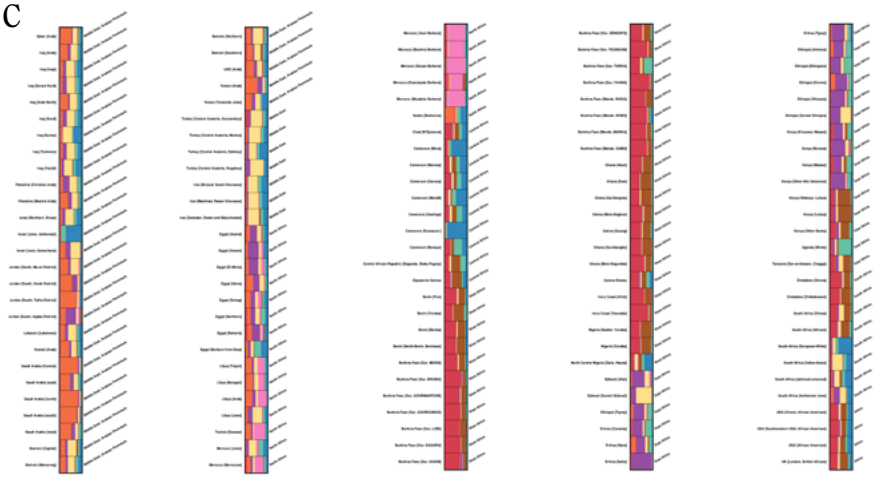
A



B



C



D

Figure 1. (A) MDS of the Middle Eastern populations (B) Phylogenetic tree of the Middle Eastern Populations (C) Q matrix of the Middle Eastern and African populations (D) Gene flow and the most dominant migration routes between Qatar and other countries.

4. Discussion

The findings of the Multidimensional Scaling (MDS) analysis indicated that the nations located within the Arabian Peninsula exhibited a distinct clustering pattern.

However, the present investigation has revealed that the Qatari demographic shares greater proximity with the Arabian Gulf countries and has formed a separate cluster when distinct from the remaining populations residing within the Middle East.

The population structure of the Y chromosome in Middle Eastern and African populations has been thoroughly analysed in the present study. Two prior investigations have explored population structure in the Middle East. The first of these studies conducted a global analysis of Y-Short Tandem Repeats (Y-STR) across 134 populations worldwide. This analysis successfully identified nine distinct clusters utilizing 19 genetic markers. However, it is noteworthy that only seven populations from the Middle East and ten from Africa were examined in comparison to the rest of the world [10].

The second study investigated 23 Middle Eastern populations, providing insights into four distinct genetic clusters through the analysis of 17 markers [11]. While both studies have made valuable contributions to our understanding of Middle Eastern population genetics, their scope was ultimately limited. These investigations may not have provided a sufficiently comprehensive or inclusive understanding of the genetic composition of Middle Eastern populations.

In the present study, we have addressed the limitations of previous investigations by conducting a more comprehensive genetic analysis of the Middle Eastern region. Our approach incorporates a broader range of populations, thereby providing a more accurate elucidation of the complex population structure in the Middle East. This approach enables us to detect subtle genetic variations and relationships that may have been overlooked in earlier studies, thus contributing to a more refined understanding of the region's genetic makeup.

The present study has yielded substantial insights into the genetic structure of Jewish populations, contributing to our understanding of their demographic history and population structure. Analysis revealed that, with the exception of the Middle Eastern Ashkenazi Jewish community, the examined Jewish populations exhibited genetic structures similar to those of non-Jewish Middle Eastern populations. This finding support the hypothesis of a shared regional ancestry among diverse Jewish communities, suggesting a common historical origin in the Middle East Furthermore, the results obtained in this study demonstrate strong agreement with previous genome-wide analyses of Jewish populations [12].

Despite encompassing two distinct geographic regions—Africa and the Middle East—and using 17 markers, the population Q-matrix produced an impressive total of 8 clusters (K=8). This finding indicates a surprisingly high number of unique geographic and population relationships. Interestingly, in a previous global

study, 9 clusters were identified using 19 markers [10]. This could be attributed to the broader scope of the survey's demographic size and ethnic composition.

The structure analysis of African populations has revealed additional substructure, elucidating distinct genetic architectures across North, East, West, and South Africa. These findings support the hypothesis of genetic substructure existing between populations associated with hunter-gatherer and agriculturalist subsistence strategies. The observed genetic differentiation aligns with both geographical distribution and historical socio-economic practices, suggesting a complex interaction between environmental adaptation and cultural factors in shaping the genetic landscape of the African continent [13, 14].

The structure of the African populations in this study showed a clusters resemblance between the four out of Africa populations and the west and south African population

Analysis of population structure revealed a notable similarity in genetic clustering patterns between the four out of Africa populations and those from West and South Africa. This observed resemblance suggests a potential shared genetic heritage or historical gene flow events connecting these geographically disparate populations.

The observed genetic similarity pattern may be attributed to the historical phenomenon of the African slave trade, one of the most significant forced migration events in human history. The transatlantic slave trade, which persisted for approximately five centuries (1400-1900 CE), can be delineated into four major phases. Of particular relevance is the final phase, which primarily originated along the West African coastline and represents the most extensive in both magnitude and duration. During this period, an estimated 12 million Africans were forcibly transported across the Atlantic Ocean, resulting in a substantial demographic shift and genetic admixture in the Americas [15].

Analysis of gene flow patterns in the Arabian Peninsula provided evidence for a model of population divergence from ancestral groups without continuous migration, which emerged as the predominant model in this investigation.

The study showed preferential migration routes, primarily along coastal areas, with a principal pathway identified from the United Arab Emirates to Kuwait and Iraq via Qatar. These results are consistent with previous research examining gene flow in the Arabian Peninsula within the context of the Out-of-Africa migration hypothesis [10]. The observed patterns support the idea of the Arabian Peninsula serving as a crucial corridor for human dispersal out of Africa [16].

5. Conclusion

Y-chromosome short tandem repeat (Y-STR) analysis of the Qatari population has emerged as a valuable molecular tool for elucidating the genetic relationships between Qatari population and other Middle Eastern ethnicities. The present study's findings underscore Qatar's integral position within the genetic landscape of the Middle East, revealing its significant contribution to the genetic composition and gene flow patterns across the Arabian Peninsula. Qatar's geographical location within the Arabian Peninsula appears to have facilitated its role as a genetic nexus, influencing the broader regional genetic make-up.

6. Acknowledgement

We would like to humbly thank Ministry of Interior of Qatar for enabling grant support for this project. We also acknowledge Qatar Defence Attaché in London, for supporting this project. We thank Promega company for providing the Kits. We wish to thank the two anonymous reviewers for their helpful comments on the manuscript.

7. Conflict of interest statement:

None

Funding

The funding was provide by Ministry of Interior of Qatar (Ref: H.R.D/S.S/D-1/1126).

Declaration of interests

None

References

[1] Qatar Statistics Authority, <https://www.psa.gov.qa/>, 2023).

[2] L. Al-Gazali, H. Hamamy, S. Al-Arrayad, Genetic disorders in the Arab world, Bmj 333(7573) (2006) 831-4.

[3] W.A. Anwar, M. Khyatti, K. Hemminki, Consanguinity and genetic diseases in North Africa and immigrants to Europe, European Journal of Public Health 24(suppl_1) (2014) 57-63.

[4] E.K. Almohammed, A. Hadi, M. Al-Asmakh, H. Lazim, The Qatari population's genetic structure and gene flow as revealed by the Y chromosome, PLoS One 18(9) (2023) e0290844.

[5] N. Takezaki, M. Nei, K. Tamura, POPTREE2: Software for Constructing Population Trees from Allele Frequency Data and Computing Other Population Statistics with Windows Interface, Molecular Biology and Evolution 27(4) (2009) 747-752.

[6] FigTree. <http://tree.bio.ed.ac.uk/software/figtree/>).

[7] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155(2) (2000) 945-59.

[8] Structure Software. <https://web.stanford.edu/group/pritchardlab/structure.html>).

[9] MIGRATE. <https://peterbeerli.com/migrate-html5/>, (accessed 2023.).

[10] H. Lazim, E.K. Almohammed, S. Hadi, J. Smith, Population genetic diversity in an Iraqi population and gene flow across the Arabian Peninsula, Scientific Reports 10(1) (2020).

[11] B.M. Albarzinji, F.M. Abdulkarim, S.A. Hussein, D. Rashid, H. Lazim, Population genetic study of 17 Y-STR Loci of the Sorani Kurds in the Province of Sulaymaniyah, Iraq, BMC Genomics 23(1) (2022) 763.

[12] D.M. Behar, B. Yunusbayev, M. Metspalu, E. Metspalu, S. Rosset, J. Parik, S. Rootsi, G. Chaubey, I. Kutuev, G. Yudkovsky, E.K. Khusnutdinova, O. Balanovsky, O. Semino, L. Pereira, D. Comas, D. Gurwitz, B. Bonne-Tamir, T. Parfitt, M.F. Hammer, K. Skorecki, R. Villems, The genome-wide structure of the Jewish people, Nature 466(7303) (2010) 238-242.

[13] M.C. Campbell, S.A. Tishkoff, African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping, Annu Rev Genomics Hum Genet 9 (2008) 403-33.

[14] F.A. Reed, S.A. Tishkoff, African human diversity, origins and migrations, Current Opinion in Genetics & Development 16(6) (2006) 597-605.

[15] G. Bertocchi, The legacies of slavery in and out of Africa, IZA Journal of Migration 5(1) (2016) 24.

[16] V. Fernandes, P. Triska, J.B. Pereira, F. Alshamali, T. Rito, A. Machado, Z. Fajkošová, B. Cavadas, V. Černý, P. Soares, M.B. Richards, L. Pereira, Genetic stratigraphy of key demographic events in Arabia, PLoS One 10(3) (2015) e0118625.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 275-280
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Mutation Rates of 23 Y-Chromosome Short Tandem Repeats (Y-STRs) in an Ecuadorian Male Population

Aníbal Gaviria[A,B+*], Santiago Cadena-Ullauri[C], Juan José Builes[D,E], Carmen Gruezo[A], Gissela Fiallos[A], Ana Karina Zambrano[C+*]

[a] *Laboratorio De Genética Molecular, Centros Médicos Especializados Cruz Roja Ecuatoriana, Quito, Ecuador.*

[b] *Hemocentro Nacional, Cruz Roja Ecuatoriana, Quito, Ecuador.*

[c] *Centro de Investigación Genética y Genómica, Facultad de Ciencias de la Salud Eugenio Espejo, Universidad UTE, Quito, Ecuador*

[d] *Laboratorio Genes SAS, Medellin, Colombia*

e Instituto de Biología, Universidad de Antioquia, Medellín, Colombia

** These authors have contributed equally to this work*

+Corresponding authors: Ulpiano Becerra. Hemocentro Nacional, Cruz Roja Ecuatoriana, Quito, Ecuador, anibalgaviria@hotmail.com, +593 2 202 0252. Mariana de Jesús and Mariscal Sucre. Centro de Investigación Genética y Genómica, Facultad de Ciencias de la Salud Eugenio Espejo, Universidad UTE, Quito, Ecuador, anazambrano17@hotmail.com, +593 2 299 0800.

Abstract

Introduction: Y-chromosome short tandem repeats (Y-STRs) are essential for forensic applications and genetic diversity studies due to their inheritance from father to son without recombination. In Ecuador, a country characterized by diverse ancestries, understanding Y-STR mutation rates is critical for enhancing genetic profiling reliability. This study aims to report the mutation rates of 23 Y-STRs in an Ecuadorian sample, addressing gaps in previous research. **Materials and Methods:** Data of a previously reported study was used, a PowerPlex®Y23 dataset of 193 father-son duos identified as mestizo. The number of mutations was assessed, and mutation rates for the 23 markers were estimated using the binomial test in RStudio v4.4.1, with confidence intervals calculated at a 95% confidence level. **Discussion:** The analysis revealed 19 mutations across the Y-STR markers, with mutation rates ranging from 0 to 0,016 for DYS391 and DYS439. These findings align with previous studies in various populations, highlighting the significance of these markers in enhancing haplotype diversity critical for population genetics and forensic analysis.

The study emphasizes the importance of interpreting mutation rates within Ecuador's demographic context, where diverse ancestries contribute to genetic variation. **Conclusion:** This study provides crucial insights into Y-STR mutation rates in the Ecuadorian mestizo population, with implications for forensic genetics, particularly in identification processes like paternity testing. Establishing reliable mutation rates is vital for accurate forensic conclusions and minimizing misinterpretations in complex genetic landscapes.

Keywords
Mutation rates, Y-chromosome, Ecuadorian population, admixed population

1. Introduction

Y-chromosome short tandem repeats (Y-STRs), located on the non-recombining portion of the Y chromosome, are commonly used in forensic applications and for studying genetic diversity within populations [1]. Since Y-STR markers are inherited from father to son without recombination, mutations are rate but provide valuable information into the genetic variation of specific populations [2].

Ecuador, has a unique demographic composition, characterized by a mixture of Native Americans, African and European ancestries, making it a critical region for studying Y chromosome diversity [3]. Reporting Y-STRs mutation rates in a diverse country such as Ecuador can improve the reliability of genetic profiling. Although previous studies have explored Y-STRs mutation rates, they often focused on smaller Ecuadorian samples or fewer loci. The present study aims to report the mutation rates of 23 Y-STRs in an Ecuadorian sample, contributing to the global understanding of Y chromosome variation and enhancing the precision of forensic and population analysis in the region.

2. Materials and methods

2.1 Data Collection

A partially anonymized dataset from Ecuadorian individuals collected by Cruz Roja Ecuatoriana and processed using PowerPlex®Y23 [Promega Corp., USA], was used to calculate the mutation rates. This dataset was previously generated for the study by Toscanini, et al. (2018) [4].

The dataset comprised a total of 193 father-son duos from Ecuadorian males auto-identified as mestizos.

2.2 Data Analysis

The number of mutations in the Y-STRs was assessed. The mutation rates for the 23 markers along with the corresponding confidence intervals (at a 95% confidence level) were estimated using the binomial test in RStudio v4.4.1[5].

32. Results

Among the 193 father-son duos analyzed, 19 mutations were identified across the 23 Y-STRs. All observed mutations involved a single mutation step change. Specifically, three mutations were detected at both DYS391 and DYS439, while two mutations were found in the markers DYS576, DYS389 II, DYS570 and DYS458. Additionally, one mutation was observed for DYS389 I, DYS448, DYS438, DYS635, and DYS456. The remaining markers, including DYS19, DYS481, DYS549, DYS533, DYS437, DYS390, DYS392, DYS643, DYS393, DYS385, and YGATAH4 did not show mutations in the population under study (Table 1).

The mutation rates for the 23 Y-STR markers in the Ecuadorian mestizo population varied from 0 for markers with no detected mutations to 0,016 for both DYS391 and DYS439.

Table 1. Mutation rates estimation for 23 Y-STRs, and corresponding confidence intervals (95%).

	Mutations	Allelic Transmission	Mutation rates	CI (95%)
DYS576	2	193	0,010	0,0013-0,0370
DYS389 I	1	193	0,005	0,00013-0,0290
DYS448	1	193	0,005	0,00013-0,0290
DYS389 II	2	193	0,010	0,0013-0,0370
DYS19	0	193	0,000	0,0000
DYS391	3	193	0,016	0,0032-0,0450
DYS481	0	193	0,000	0,0000
DYS549	0	193	0,000	0,0000
DYS533	0	193	0,000	0,0000
DYS438	1	193	0,005	0,00013-0,0290
DYS437	0	193	0,000	0,0000
DYS570	2	193	0,010	0,0013-0,0370
DYS635	1	193	0,005	0,00013-0,0290

DYS390	0	193	0,000	0,0000
DYS439	3	193	0,016	0,0032-0,0450
DYS392	0	193	0,000	0,0000
DYS643	0	193	0,000	0,0000
DYS393	0	193	0,000	0,0000
DYS458	2	193	0,010	0,0013-0,0370
DYS385	0	193	0,000	0,0000
DYS456	1	193	0,005	0,00013-0,0290
YGATAH4	0	193	0,000	0,0000
TOTAL	19	4246		

4. Discussion

The mutation rates observed in this study for the 23 Y-STR markers in the Ecuadorian mestizo population contribute valuable data to the existing literature on Y-chromosome diversity. The mutation rates ranged from 0 in eleven of the markers to a maximum of 0,016 for both DYS391 and DYS439. These findings are consistent with previous studies conducted on other population, such as those in Argentina, Brazil, Colombia, Portugal, Spain and Venezuela, where they found 9 mutations for DYS391 and 8 mutations for DYS439 [6]. The high mutation rates in these markers highlight the role increasing the haplotype diversity within the Y-STR marker set, which is critical for applications in population genetics and forensic analysis [6, 7]. These markers, due to their higher mutation rates, require carful interpretation in forensic exclusion scenarios to minimize the risk of misinterpreting genetic evidence, particularly in highly admixed population like Ecuador's [8].

Furthermore, the results of this study emphasize the importance of contextualizing Y-STRs mutation data within Ecuador's demographic history, which is characterized by a mixture of native American, European and African ancestries [3]. By establishing precise mutation rates for this population, this study increases the accuracy of genetic profiling when combined with autosomal markers, offering improved reliability for both forensic investigations and population genetics studies [9]. The integration of such population-specific data is vital for refining exclusion criteria and ensuring that forensic conclusions are drawn with greater confidence and accuracy in diverse genetic landscapes.

5. Conclusion

This study reports the mutation rates of 23 Y-STR markers in the Ecuadorian mestizo population, contributing to the understanding of Y-chromosome diversity within an admixed population. The data generated in this research have significant implication, especially for forensic genetics, as Y-STR markers are useful in identification processes such as paternity testing and criminal implications. The study highlights the importance of accurately determining mutation rates, especially for markers with higher mutation frequencies, as these can critically influence the interpretation of genetic evidence. In forensic analysis, the ability to establish clear exclusion criteria based on reliable mutation rates is vital to ensure accurate identification and minimize the risk of misinterpretation. Overall, this study contributes to the global understanding of Y-STR mutation dynamics, emphasizing the need for population specific frequency data for genetic analysis.

6. Acknowledgements

The study received the support of Cruz Roja Ecuatoriana. We extend our gratitude to the Universidad UTE for their support.

7. Conflict of interest statement

The authors declare no conflicts of interest.

8. References

[1] Gaviria A, Sánchez ME, Morejón G, et al. Characterization and Haplotype analysis of 11 Y-STR loci in Ecuadorian population. *Forensic Sci Int Genet Suppl Ser* 2013; 4: e310–e311.

[2] Ballantyne KN, Goedbloed M, Fang R, et al. Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications. *Am J Hum Genet* 2010; 87: 341.

[3] Zambrano AK, Gaviria A, Cobos-Navarrete S, et al. The three-hybrid genetic composition of an Ecuadorian population using AIMs-InDels compared with

autosomes, mitochondrial DNA and Y chromosome data. *Sci Rep*; 9. Epub ahead of print 2019. DOI: 10.1038/s41598-019-45723-w.

[4] Toscanini U, Gaviria A, Pardo-Seco J, et al. The geographic mosaic of Ecuadorian Y-chromosome ancestry. *Forensic Sci Int Genet* 2018; 33: 59–65.

[5] RStudio Core Team. RStudio: Open source and enterprise-ready software for R. *RStudio Home*, https://www.rstudio.com/ (2018, accessed 10 January 2018).

[6] Gusmão L, Sánchez-Diz P, Calafell F, et al. Mutation rates at Y chromosome specific microsatellites. *Hum Mutat* 2005; 26: 520–528.

[7] Alves C, Gusmão L, Barbosa J, et al. Evaluating the informative power of Y-STRs: A comparative study using European and new African haplotype data. *Forensic Sci Int* 2003; 134: 126–133.

[8] Jobling MA, Gill P. Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics* 2004; 5: 739–751.

[9] Gaviria A, Vela M, Fiallos G, et al. Mutation rates for 29 short tandem repeat loci from the Ecuadorian population. *Forensic Sci Int Genet Suppl Ser* 2017; 6: e229–e230.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 281-288
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Evaluation of mutation rates in Nigerian father-son pairs using Y-STR markers

Simon Knights[a], Sharlize Pedroza-Matute[a,b], Thomas Haizel[a], Oyenike Arike Adeyemo[c], Syed Sibte Hadi[d], Sasitaran Iyavoo[a,b*]

[a] *AttoLife Limited, Norwich, United Kingdom.*

[b] *University of Lincoln, Lincoln, United Kingdom.*

[c] *University of Lagos, Lagos, Nigeria.*

[d] *Naif Arab University for Security Sciences, Riyadh, Saudi Arabia.*

** Corresponding author at: School of Natural Sciences, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, United Kingdom, siyavoo@lincoln.ac.uk.*

Abstract

Male-specific Y-STR markers are a valuable complement to autosomal STR testing, especially in sexual assault cases, but mutation rates can affect interpretations and require careful evaluation. The Yfiler™ Plus PCR Amplification Kit, containing 27 Y-STR loci, including 7 rapidly mutating loci (mutation rates > 1x10-2), was assessed using Nigerian samples. Buccal swabs from 148 father-son pairs underwent DNA extraction with the QIAamp® DNA Mini Kit. Paternity was confirmed using the VeriFiler™ Express PCR Amplification Kit with a Combined Paternity Index threshold of 10,000. All 148 pairs showed no paternity mismatches. Y-STR profiles were generated with the Yfiler™ Plus kit, and mutation rates with binomial 95% confidence interval (CI) were calculated. Results were compared with mutation rates from the Y Chromosome Haplotype Reference Database (YHRD). A total of 29 one-step mutations were identified across 15 markers, with the highest mutation rates at DYS518 (0.0338, 95% CI: 0.0111-0.0771), DYS458, and DYS449 (0.0270, 95% CI: 0.0074-0.0678). Approximately 48% of mutations were repeat gains, and 52% were losses. Except for three pairs with two mutations, all mutations appeared once per pair. YHRD mutation rates for DYS458 and DYS518 were lower than the 95% CI obtained in the Nigerian dataset. Additionally, the rapidly mutating markers; DYS627, DYS570, and DYF387S1 had mutation rates lower than the 0.01 rate indicated by the manufacturer. Although limited by the small sample size, this study highlights the need to expand Y-STR mutation rate evaluations and establish population-specific rates for accurate forensic and relationship testing.

Keywords

Yfiler™ Plus PCR Amplification Kit, Y-STRs, rapidly mutating markers, mutation rates, father-son pairs, population-specific genetics.

Introduction

The Y chromosome is a male lineage marker that is transmitted almost intact from father to son across generations. Variations arise through recombination with the X chromosome, which occurs only in two pseudo-autosomal regions outside of the male-specific, non-recombining region, and through mutations. Due to the high number of mitotic events in spermatogenesis, mutations between father and son are expected. However, the mutation rate has been observed to vary dependent on the population and the region of the chromosome in which the genetic marker is located [1].

The characteristics of the male-specific part of the Y chromosome, inherited as a single set of genetic markers (i.e., haplotype), make it ideal for use in various forensic scenarios. Analysis of Y-chromosome Short Tandem Repeat (Y-STR) markers, for example, can enable the identification of Y haplotypes specific to the paternal lineage of a male perpetrator in sexual assault cases. Y-STRs can also be utilised in parentage and kinship testing, as well as in identification of missing persons and mass disaster victims. Due to their high inter-population variability, Y-STRs can additionally be employed to infer biogeographical origins, which can be valuable as an investigative lead for unknown individuals and for studying historical migration patterns [2-3].

Among the commercial kits available for Y-STR typing, the Yfiler™ Plus PCR Amplification Kit contains 27 Y-STR loci, including 7 rapidly mutating loci (DYS576, DYS627, DYS518, DYS570, DYS449, and DYF387S1a/b) with mutation rates reportedly exceeding $1x10^{-2}$ [4-5]. This kit incorporates the original 17 loci from the Yfiler™ PCR Amplification Kit, along with an additional 10 loci, making it one of the most comprehensive options available for forensic and relationship testing involving Y-chromosome analysis.

Forensic applications of Y-STR typing require information on haplotype frequencies worldwide. To facilitate this, the Y Chromosome Haplotype Reference Database (YHRD) was established to generate frequency estimates for quantitative assessments. The database contains extensive data on different populations and also provides combined mutation rates for common Y-STRs [6]. Although the YHRD website includes a national database for Nigeria, with haplotype counts of

497 for Y17 and 358 for Y27 (Release 69), and other studies have been published on forensic parameters for Y-STRs in Nigerian populations [7-8], specific data on mutation rates for Nigerians are still lacking.

Given the impact that mutations can have on the interpretation of forensic cases and the differences observed across population, the objective of this study was to evaluate the mutation rate of Y-STRs present in the Yfiler™ Plus kit in Nigerian father-son pairs.

Material studied, methods, techniques

Study samples

After obtaining informed consent, buccal swabs were collected from 148 father-son pairs with self-declared Nigerian ethnicity.

DNA extraction and Y-STR typing

DNA was extracted using the QIAamp® DNA Mini Kit (Qiagen) and amplified with the Yfiler™ Plus *PCR Amplification Kit (Applied Biosystems™)*. The resulting profiles were manually reviewed to identify mutations within each father-son pair. These mutations were categorised into repeat gains or losses. The mutation rates were calculated by dividing the number of observed mutations for each marker by the total number of meiotic transfers observed. The binomial 95% confidence intervals (CI) of the mutation rates were calculated as previously reported [9]. The results were then compared with mutation rates reported on the YHRD website [6] (Release 69).

Confirmation of relationships

Paternity of each father-son pair was confirmed using the VeriFiler™ Express PCR Amplification Kit *(Applied Biosystems™)* [10], applying a Combined Paternity Index (CPI) threshold of 10,000.

Results

Using the results from the VeriFiler™ Express kit, paternity was confirmed for all 148 father-son pairs, with no mismatches observed. The Yfiler™ Plus Y-STRs typing results revealed a total of 29 one-step mutations across 15 markers among all

the pairs under investigation. As shown in Figure 1, eight of these loci (53%) had only one mutation, while DYS518 emerged as the locus with the highest number of mutations, followed by DYS458 and DYS449.

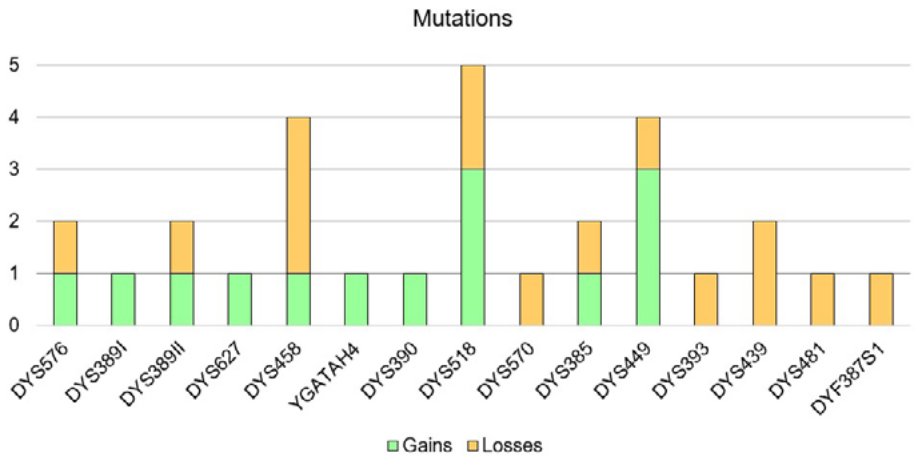


Figure 1: Distribution of the 29 mutations observed in the father-son pairs analysed in this study.

Overall, the mutations appeared in 26 father-son pairs. Each mutation occurred once in different pairs, except in three pairs where two mutations were observed (Figure 2a). Most of the mutations were repeat losses (52%), while the remaining 48% were repeat gains (Figure 2b).

a)

F-S Pair	F-S Alleles	Marker	F-S Pair	F-S Alleles	Marker
A	18	DYS576	P	21	DYS570
	19			20	
B	18		Q	17,18	DYS385
	17			16,18	
C	13	DYS389I	R	17,17	
	14			17,18	
C	30	DYS389II	S	31	DYS449
	31			32	
D	31		T	33	
	30			34	
E	19	DYS627	U	29	
	20			30	
F	19	DYS458	V	29	
	18			28	
G	17		W	14	DYS393
	16			13	
H	18		X	13	DYS439
	19			12	
D	19		Y	14	
	18			13	
I	12	YGATAH4	H	25	DYS481
	13			24	
J	22	DYS390	Z	39,41	DYF387S1
	23			39,40	
K	42	DYS518	b)		
	43				
L	41				
	40				
M	37				
	36				
N	39				
	40				
O	40				
	41				

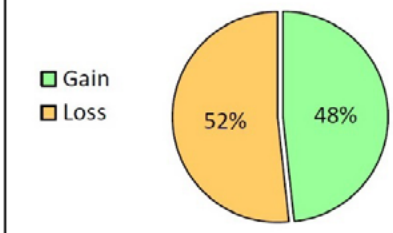


Figure 2: a) Alleles observed in father-son (F-S) pairs exhibiting mutations. For each pair, the father's alleles are shown at the top, and the son's alleles at the bottom. Gain mutations are highlighted in green, and loss mutations in orange. b) Pie chart representing the percentage of gain and loss mutations.

Consistent with the observations in Figure 1, the highest mutation rates were found at loci DYS518 (0.0338, 95% CI: 0.0111-0.0771), DYS458 and DYS449 (0.0270, 95% CI: 0.0074-0.0678) (Table 1). When compared to the combined mutation rates reported in the YHRD database, the mutation rates for DYS458 and DYS518 were found to be lower than the 95% CI obtained in this study. Additionally, in this study, the rapidly mutating markers DYS627, DYS570, and DYF387S1 exhibited mutation rates lower than the 0.01 rate indicated by the kit manufacturer.

Table 1: Mutation rate and confidence interval obtained from the study samples for each locus with mutations. The last column shows the combined mutation rates reported in the YHRD database. The loci in bold represent the rapidly mutating markers.

Yfiler™ Plus locus	Mutation rate	Binomial 95% confidence interval	YHRD mutation rate
DYS576	0.0135	0.0016-0.0480	0.0116
DYS389I	0.0068	0.0002-0.0371	0.0024
DYS389II	0.0135	0.0016-0.0480	0.0050
DYS627	0.0068	0.0002-0.0371	0.0139
DYS458	0.0270	0.0074-0.0678	0.0065
YGATAH4	0.0068	0.0002-0.0371	0.0026
DYS390	0.0068	0.0002-0.0371	0.0020
DYS518	0.0338	0.0111-0.0771	0.0106
DYS570	0.0068	0.0002-0.0371	0.0080
DYS385	0.0135	0.0016-0.0480	0.0028
DYS449	0.0270	0.0074-0.0678	0.0096
DYS393	0.0068	0.0002-0.0371	0.0013
DYS439	0.0135	0.0016-0.0480	0.0045
DYS481	0.0068	0.0002-0.0371	0.0038
DYF387S1	0.0068	0.0002-0.0371	0.0065
Overall	0.0131	0.0088-0.0187	

Discussion

Mutations can have a significant impact on the interpretation of forensic evidence. A high mutation rate, for example, can lead to false exclusions in paternity testing cases and can affect the identification of mass disaster victims [11]. For this reason, using accurate mutation rate estimates is essential to avoid misinterpretations. The present study indicates that relying on the combined Y-STRs mutation rates provided in the YHRD database may be inaccurate for the Nigerian population. In fact, some rapidly mutating markers exhibited lower mutation rates than expected, while other loci showed increased mutation rates. However, this study was limited by the small number of father-son pairs analysed, highlighting the need for further efforts to establish population-specific databases for Nigerian and other understudied populations, using larger and more diverse samples. This applies not only to Y-STR frequencies but also to mutation rates. Such databases will have significant implications for forensic casework, paternity testing, and population genetics, particularly given that mutation rates can influence the interpretation of forensic evidence.

Conclusion

The results of this study, including the lower-than-expected mutation rates for certain rapidly mutating markers and vice versa, suggest that evaluations of Y-STR mutation rates should be expanded. Additionally, the importance of establishing population-specific mutation rates is emphasised. Greater knowledge of Y-STR mutation rates would enhance the use of the Yfiler™ Plus kit, or any other Y-STR kits, in forensic and relationship testing, thereby improving accuracy and reducing the likelihood of errors.

Conflict of interest statement

All claims expressed in this article are solely those of the authors and do not necessarily represent the views of their affiliated company or organisation.

References

[1] D. Syndercombe Court, The Y chromosome and its use in forensic DNA analysis, Emerg. Top. Life Sci. 5 (2021) 427-441.

[2] M. Kayser, Forensic use of Y-chromosome DNA: a general overview, Hum. Genet. 136 (2017) 621–635.

[3] L. Roewer, Y-chromosome short tandem repeats in forensics—Sexing, profiling, and matching male DNA, WIREs Forensic Sci. 1 (2019) e1336.

[4] S. Gopinath, C. Zhong, V. Nguyen, et al., Developmental validation of the Yfiler® Plus PCR Amplification Kit: An enhanced Y-STR multiplex for casework and database applications, Forensic Sci. Int. Genet. 24 (2016) 164-175.

[5] K.N. Ballantyne, V. Keerl, A. Wollstein, et al., A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages, Forensic Sci. Int. Genet. 6 (2012) 208-218.

[6] S. Willuweit, L. Roewer, The new Y Chromosome Haplotype Reference Database, Forensic Sci. Int. Genet. 15 (2015) 43-48.

[7] S.T. Fakorede, K.O. Adekoya, O.O. Iroanya, et al., Forensic parameters of 10 Y-STR markers in the Nigerian Hausa, Igbo, and Yoruba populations, UMYU Scientifica, 3 (2024) 103-112.

[8] B. Martinez, L. Catelli, M. Romero, et al., Forensic evaluation of 27 y-str haplotypes in a population sample from Nigeria, Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) E289-E291.

[9] M. Goedbloed, M. Vermeulen, R.N. Fang, et al., Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR® Yfiler® PCR amplification kit, Int. J. Leg. Med. 123 (2009) 471-482.

[10] J. Perry, T. Munshi, T. Haizel, et al., Validation of reduced volume VeriFiler™ Express PCR Amplification Kit for buccal swab samples extracted using Prep-n-Go™ Buffer. J. Forensic Sci. 67 (2022) 1971-1978.

[11] J.M. Butler, Chapter 6 – Biology of STRs: stutter products, non-template addition, microvariants, null alleles and mutation rate. In: J.M. Butler, Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers, Second Edition, Elsevier Academic Press, New York, 2005, pp. 123-144.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 289-295
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Genetic characterization of Y-chromosomal STRs in Santander department of Colombia

A. Castillo[1], Z. Köksal[2], G. Mantilla[1], C. I. Vargas[1], H. Ossa[3], L. Gusmão[4]

[1]*Laboratorio de Genética, Escuela de Medicina, Universidad Industrial de Santander (UIS), Bucaramanga, Colombia.*

[2]*Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.*

[3]*Laboratorio de Genética y Biología Molecular, Bogotá, Colombia.*

[4]*DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro, Rio de Janeiro, Brazil.*

Corresponding author: Adriana Castillo · Tel: (+57) 3004947194 Fax:(+57) 6076356641

E-mail: castillo@uis.edu.co

Abstract

The use of uniparental markers such as those located in the non-recombining region of the Y chromosome can be very useful to solve some forensic cases. Namely, Y-STRs are relevant in investigations of sexual assaults since autosomal markers can be difficult to analyze in evidence with mixed contributions from a female victim and a male aggressor. Despite its importance, the use of Y-chromosomal markers in Colombia is limited. Colombia is an admixed population with a high ancestry variation, where the Native, European, and African contributions vary widely among population from different regions. Nevertheless, there is a lack of Y-STR haplotypic databases representative of most country regions. Moreover, the few haplotypic profiles in the YHRD only include 17 Y-STRs, with no data available for the most recent kits. These population databases are important for understanding the genetic diversity of the country and its substructure, as well as to estimate the Y-STR haplotype frequencies in each region. The aim of this study was to contribute with data from 97 unrelated individuals born in the department of Santander (northeast Colombia), genotyped with the PowerPlex Y23. A total of 95 haplotypes were detected, for a haplotype diversity of 0.9998 ± 0.1500. The classification of paternal lineages based on haplogroups showed a predominant European contribution (85.6 %), followed by Amerindian (13.4 %) and African (1.0 %) contributions.

Keywords

Y chromosome; forensic genetics; paternal ancestry; male lineages.

1. Introduction

The analysis of genetic markers located in the non-recombinant region of the Y chromosome is very useful to solve forensic cases as a complementary analysis of autosomal STRs. Namely, due to its male specificity and uniparental inheritance, Y-chromosomal markers can be useful (i) in situations involving samples from sexual assaults containing mixed profiles, (ii) to exclude male suspects involved in a crime, or (iii) in kinship cases, when only patrilineal relatives are available [1]. The most used Y-chromosomal markers are the Short Tandem Repeats (Y-STRs) because they have a high mutation rate, which results in a great variability of highly discriminative haplotypes in the populations [2].

The genotyping of Y-STRs by capillary electrophoresis remains the gold standard in forensic genetics [3]. Since its use as a forensic tool, the development of commercial kits with an increasing number of Y-STRs genotyped in a single multiplex reaction has been observed. These extended sets of markers allow for greater genetic diversity of the haplotypes found in a population and, therefore, increase the discrimination power that can be achieved from their analysis [3,4].

The genetic variability present in different populations makes it necessary to determine the haplotype distribution existing in each population, with the creation of specific databases to be used in solving forensic cases [5].

Colombia is a country located in the northwestern corner of South America, made up of 32 departments, including Santander, located in the northeast of the country. Analysis of autosomal genetic markers has shown that the country's current population resulted from the admixture of three continental groups [6,7]. Admixture processes occurred among previously existing Native Americans, with Europeans who arrived with the conquest in the 15th century, and with Africans brought as slaves in the 16th century. However, these studies have revealed that the percentage of intercontinental admixture varies across the country's regions. Therefore, it is necessary to analyze each department and build specific databases that allow an objective evaluation of forensic evidence. Although studies with Y-chromosomal STRs have been conducted in Colombia, these have included up to 17 markers, only from populations in some departments [8,9]. This study presents the results of the analysis of 23 Y-STRs that will form a database for forensic use in Santander (Colombia).

2. Material and methods

2.1. Samples

Blood samples from 97 unrelated male individuals born in the department of Santander were collected in EDTA tubes. An informed consent was signed by all participates and this study was approved by the ethics committee of the Universidad Industrial de Santander (CEINCI - 11/08/2023). DNA was extracted with the PureLink genomic DNA kit (Invitrogen). Subsequently, 23 Y-STRs included in the PowerPlex® Y23 System (Promega) were genotyped, following the manufacturer's protocol. PCR products were separated and detected by capillary electrophoresis on an ABI 3500 genetic analyzer (Applied Biosystems). The results were analyzed using the GeneMapper™ Software v5.0 (Applied Biosystems). Haplotypes were assigned following the ISFG recommendations [10].

2.2. Statistical analyses

Allelic, and haplotype/haplogroup frequencies were calculated by direct counting. The Arlequin software v3.5.2.2 was used to calculate haplotype diversity (H). The random match probability (MP) and discrimination capacity (DC) were calculated according to Purps [4]. Haplogroup predictions based on Y-STR haplotypes' data were performed using the NEVGEN tool (https://www.nevgen.org/, accessed on 20 March 2024).

3. Results and Discussion

3.1. Single-locus analysis

A total of 158 different alleles were detected in the 23 Y-STRs genotyped in 97 individuals, with an average of seven alleles per marker, ranging from 3 alleles (at DYS389I and DYS393) to 10 alleles at DYS481 (Table 1). At DYS385, a total of 33 haplotypic combinations were detected (data not shown), with haplotype "11,14" showing the highest frequency (32%).

Table 1. Allelic frequencies for the 23 Y-STRs

Allele	DYS576	DYS389I	DYS448	DYS389II	DYS19	DYS391	DYS481	DYS549	DYS533	DYS438	DYS437	DYS570	DYS635	DYS390	DYS439	DYS392	DYS643	DYS393	DYS458	DYS456	YGATAH4
6						0.010															
7										0.010											
8																	0.021				
9						0.041			0.031	0.124							0.093				
10						0.495		0.021	0.041	0.175					0.021		0.577				0.010
11		0.113				0.433		0.052	0.216	0.165					0.268	0.247	0.155				0.392
12		0.639				0.021		0.320	0.639	0.513					0.536	0.052	0.113	0.196		0.010	0.536
13		0.247			0.196			0.485	0.051	0.010					0.155	0.526	0.031	0.680		0.010	0.062
14	0.021				0.567			0.124	0.021		0.381	0.010			0.021	0.144	0.010	0.124	0.021	0.041	
15	0.052			0.062	0.196						0.526	0.031				0.021			0.124	0.545	
16	0.124			0.619	0.041						0.093	0.113				0.010			0.289	0.268	
17	0.268			0.227								0.454							0.309	0.103	
17.2																			0.010		
18	0.268		0.134	0.072			0.010					0.237							0.175	0.021	
18.2																			0.010		
19	0.227		0.454	0.021								0.103							0.052		
20	0.041		0.186				0.031					0.031	0.010						0.010		
21			0.155				0.082					0.013	0.216	0.041							
22			0.052				0.402					0.010	0.113	0.062							
23			0.021				0.175						0.546	0.216							
24							0.175						0.093	0.505							
25							0.052						0.010	0.165							
26							0.031						0.010	0.010							
27							0.031														
29							0.010														

Intermediate alleles were detected at 2 of the 23 loci (DYS458 and DYS385) and no copy number variants were found. The range of alleles detected agree with that found by Purps et al. [4]. However, the mean number of alleles per marker was lower, as well as the number of intermediate alleles, as expected considering the large difference in the number of samples/populations analyzed.

3.2 Forensic parameters

In our sample, 95 unique haplotypes (97.9%) were found and only one haplotype was shared by two individuals. Both MP and DC values were high (1 in 5000 and 0.9794, respectively). As expected, these values are higher than those previously reported for Santander for 11 Y-STRs [8], and for Valle, Cauca and Nariño analyzed for 17 Y-STRs [9], emphasizing the importance of enlarging the set of Y-STRs to increase discrimination power. The H, as well as the MP found were very similar to those reported in other studies carried out in Latin American populations and the discrimination capacity was higher except for the Mexican one, despite having the smallest sample size. (Table 2).

Table 2. Diversity and forensic parameters for PPY23.

Population	N	K	S	H	MP	DC
This study	97	96	95	0.9998 ± 0.1500	1 in 5000	0.9794
Argentina [4]	621	599	579	0.9999 ± 0.0001	1 in 10000	0.9324
Brazil [5]	1319	1231	1230	0.9998 ± 0.0005	1 in 5000	0.9325
Ecuador [9]	269	264	258	0.9998 ± 0.0000	1 in 5000	0.9591
México [10]	307	255	253	-----	---	0.9870
Paraguay [2]	537	480	441	0.9994 ± 0.0002	1 in 1667	0.8212

N: Sample size; k: Number of different haplotypes; S: Singletons; H: Haplotype diversity; MP: Match probability; DC: Discrimination capacity (S/N).

3.3. Inferred ancestry of Y-chromosomal haplotypes

The haplogroups of the 97 samples were assigned using the NEVGEN prediction tool. A total of 83 samples were assigned to haplogroups of European ancestry (85.6%), followed by 13 samples with Native American haplogroups (13.4%) and one sample belonging to an African haplogroup (1%). The high frequency of European paternal ancestry corroborates the sex-biased admixture pattern between the Colombian ancestors, predominantly European men with Native American women [6,7,18]. These results agree with those from other studies carried out in admixed populations from South America [5,14,16,17], except for the admixed Mexican population, where the Native American haplogroups were the most frequent [15]. Finally, both in our population and in other South American populations that have been studied so far, African haplogroups were the least frequent [14,15,16,17].

4. Conclusions

This study provides 97 haplotypes from the department of Santander, genotyped for 23 Y-STRs, which are the only data available for this set of markers in the Colombian population. This work allowed to increase the number of haplotypes available for the department of Santander, which together with samples previously typed for 11 Y-STRs, make up a total of 449 haplotypes, thus improving the detection capacity of the high diversity of haplotypes, and increasing the power of discrimination.

Conflict of interest

None

Financial support

AC was supported by Vicerrectoria de Investigación y Extension (VIE) - Universidad Industrial de Santander, Proyecto 3938-2023. LG is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (ref. 315007/2023-0) and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ (CNE-2022 and project ref. E-26/211.369/2021).

References

[1] M. Kayser, Forensic use of Y-chromosome DNA: a general overview, Hum Genet (2017) 136:621–635. https://doi.org/10.1007/s00439-017-1776-9

[2] J. Ribeiro, M. Romero, F. Simão, et al., Analysis of 23 Y-STRs in a population sample from eastern Paraguay, Forensic Science International: Genetics, 37 (2018) e20-e22. https://doi.org/10.1016/j.fsigen.2018.08.007

[3] G. Soldati, S. Turrina, M. Treccani, et al., Concordance study on Y-STRs typing between SeqStudio™ genetic analyzer for HID and MiSeq™ FGx forensic genomics system, Molecular Biology Reports. (2023) 50:9779–9789. https://doi.org/10.1007/s11033-023-08808-4

[4] J. Purps, S. Siegert, S. Willuweit, et al., A global análisis of Y-chromosomal haplotype diversity for 23 STR loci, Forensic Science International: Genetics. 12 (2014) 12-23. http://dx.doi.org/10.1016/j.fsigen.2015.12.005

[5] U. Toscanini, F. Brisighelli, F. Moreno et al., Analysis of Y-chromosome STRs in Chile confirms an extensive introgression of European male lineages in urban populations, Forensic Science International: Genetics. 21 (2016) 76-80. http://dx.doi.org/10.1016/j.fsigen.2015.12.005

[6] W. Rojas, M.V. Parra, O. Campo et al., Genetic make up and structure of Colombian populations by means of uniparental and biparental DNA markers, Am J Phys Anthropol. 143 (2010) 13-20. https://doi.org/10.1002/ajpa.21270

[7] H. Ossa, J. Aquino, R. Pereira, et al., Outlining the Ancestry Landscape of Colombian Admixed Populations, PLoSONE 11 (2016), e0164414. https://doi.org/10.1371/journal.pone.0164414

[8] A. Castillo, K. Rueda, A. Pico et al. Comparing different population groups in Santander—Colombia through Y-STR haplotype análisis, Forensic Science International: Genetics Supplement Series 5 (2015) e482–e483. http://dx.doi.org/10.1016/j.fsigss.2015.09.191

[9] S. Ávila, I. Briceño. A. Gómez, Genetic population analysis of 17 Y-chromosomal STRs in three states (Valle del Cauca, Cauca and Nariño) from Southwestern Colombia, Journal of Forensic and Legal Medicine, 16 (2009) 204–211. https://doi.org/10.1016/j.jflm.2008.12.002

[10] L. Gusmão, J.M. Butler, A. Carracedo, P. Gill, M. Kayser, W.R. Mayr, et al., DNA-commission of the international society of forensic genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis, Forensic Sci. Int. 157 (2006) 187–197. https://doi.org/10.1016/j.forsciint.2005.04.002

[11] U. Toscanini, C. Vullo, G. Berardi, et al. A comprehensive Y-STR portrait of Argentinean populations, Forensic Science International: Genetics, 20 (2026) 1 – 5. http://dx.doi.org/10.1016/j.fsigen.2015.09.002

[12] S.T. Ferreira, K.A. Paula, F.A. Maia,et al. Routine analysis of sexual assault cases in Brasília, Brazil, using 23 Y chromosomal markers, Forensic Sci. Int.:Genet. Suppl. Ser. 5 (2015) e619–e621. https://doi.org/10.1016/j.fsigss.2015.09

[13] I. Ambrosio, D. Braganholi, L. Orlando et al., Mutational data and population profiling of 23 Y-STRs in three Brazilian populations, Forensic Science International: Genetics, 48 (2020) 102348. https://doi.org/10.1016/j.fsigen.2020.102348

[14] U. Toscanini, A. Gaviria, J. Pardo-Seco, et al., The geographic mosaic of Ecuadorian Y-chromosome ancestry, Forensic Science International: Genetics, 33 (2018) 59-65. https://doi.org/10.1016/j.fsigen.2017.11.011

[15] Y. L. López-Ramírez, J. A. Aguilar-Velázquez, M. López-Armenta et al. Paternal lineages and forensic parameters based on 23 Y-STRs (Powerplex® Y23) in Mestizo males from Mexico City, International Journal of Legal Medicine (2020) 134:199–202 https://doi.org/10.1007/s00414-019-02183-1

[16] P. Rodrigues, I. F. Velázquez, J. Ribeiro, et al. Tierra Del Fuego: What Is Left from the Precolonial Male Lineages? Genes 2022, 13, 1712. https://doi.org/10.3390/genes13101712

[17] F. Simão, J. Ribeiro, C. Vullo, et al. The Ancestry of Eastern Paraguay: A Typical South American Profile with a Unique Pattern of Admixture, Genes (2021) 12:1788. https://doi.org/10.3390/genes12111788

[18] A. Castillo, F. Rondón, G. Mantilla, et al. Maternal ancestry and lineages diversity of the Santander population from Colombia, Forensic Sci Res. (2023) 21;8(3):241-248. https://doi.org/10.1093/fsr/owad032

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 297-303
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Exploring Y-DNA haplogroup diversity in the British Isles

Sasitaran Iyavoo[a,b*], Sharlize Pedroza-Matute[a,b], Thomas Haizel[b]

[a] *University of Lincoln, Lincoln, United Kingdom.*
[b] *AttoLife Limited, Norwich, United Kingdom.*

** Corresponding author at: School of Natural Sciences, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, United Kingdom, siyavoo@lincoln.ac.uk.*

Abstract

The Yfiler™ Plus PCR Amplification Kit is a Y-chromosome STR kit designed for the specific amplification of male DNA samples. This makes it particularly useful in sexual assault cases and paternal lineage studies. Additionally, the high inter-population variability of Y-STRs makes the kit ideal for investigating biogeographical differences and migration patterns. The aim of this study was to investigate the haplogroups of the English, Irish, Welsh, and Scottish populations to support the development of population-specific Y-STR databases and advance our understanding of the genetics of the British Isles. Buccal swabs were collected from 888 unrelated male individuals: 461 English, 191 Irish, 124 Welsh, and 112 Scottish. DNA was extracted using Prep-n-Go™ Buffer and amplified with the Yfiler™ Plus kit. The resulting data passed the Y Chromosome Haplotype Reference Database (YHRD) quality checks and were uploaded to the database. Haplogroups were identified using the Y-DNA Haplogroup Predictor - NevGen online tool, and haplogroup diversity was calculated. A total of 46 haplogroups were observed, with R1b M269 being the most common haplogroup across all populations (64% in English, 85% in Irish, 92% in Welsh, and 73% in Scottish). I1 was the second most common haplogroup in both the English (13%) and Scottish (6%) populations, while I2a2a M223 was the second most common in the Welsh (2%) and Irish (6%) populations. R1a M198 and E1b1b > V13 were the other two most frequently observed haplogroups, while the remaining 41 haplogroups were found five or fewer times across all populations. Although the Scottish population showed a higher number of haplogroups relative to the sample size tested, greater diversity was observed in the English (0.57) and Scottish (0.46) populations compared to the Irish (0.28) and Welsh (0.14) populations. This study highlights genetic differences between the four populations, with the distribution of haplogroups supporting previous findings. However, further studies with more markers and larger sample sizes are needed to continue populating Y-marker databases.

Keywords
Yfiler™ Plus PCR Amplification Kit, Y-STRs, Y haplogroups, population genetics, British Isles, haplogroup diversity.

Introduction

The Yfiler™ Plus PCR Amplification Kit is a multiplex kit that detects 27 Y-chromosome Short Tandem Repeat (Y-STR) loci [1]. As the Y chromosome is passed from father to son across generations and is absent in females, it is particularly useful in several forensic applications. For example, in sexual assault cases, Y-STR testing can isolate male DNA, identifying the paternal lineage of the perpetrator. It is also valuable in relationship testing and in the identification of missing persons and victims of mass disasters.

The male-specific region of the Y chromosome is inherited as a linkage block without recombination. However, mutations occurring during male gamete production accumulate across generations, leading to significant inter-population variability. These features make the study of Y markers ideal for investigating biogeographical differences and migration patterns [2-3].

Haplogroups, which represent shared paternal lineage and common ancestry, are defined by specific markers and named using an alphanumeric notation system. They are classified in phylogenetic trees, which are used for demographic studies. Y haplogroups are typically defined by Single Nucleotide Polymorphisms (SNPs), while Y-STR linkage groups are referred to as haplotypes [4]. Several Y-haplogroup predictor tools have been developed to infer haplogroups from Y-STR profiles [5].

Since haplotypes and haplogroups are characterised by multiple linked markers, their population frequencies cannot be derived from individual allelic frequencies using the product rule. Therefore, it is essential to continually populate databases with data from various populations to identify rare profiles and support accurate statistical calculations [6].

This study aimed to investigate the Y haplogroups of English, Irish, Welsh, and Scottish populations to aid in the development of population-specific Y-STR databases and enhance understanding of the genetic structure of the British Isles.

Material studied, methods, techniques

Study samples

Following the collection of informed consent, buccal swabs were taken from 888 unrelated male individuals who self-identified their ethnicity as being from various regions of the British Isles: 461 English, 191 Irish, 124 Welsh, and 112 Scottish.

DNA extraction and Y-STR typing

DNA was extracted using the Prep-n-Go™ Buffer (Applied Biosystems™) [7] and amplified with the Yfiler™ Plus PCR Amplification Kit (Applied Biosystems™), following the manufacturer's instructions. Capillary electrophoresis was performed on the ABI Prism® 3500xL Genetic Analyzer (Applied Biosystems™), and profiles were analysed using GeneMapper® ID-X v1.6 software (Applied Biosystems™).

Data analysis

All obtained Y-STR profiles passed the quality checks of the Y-Chromosome STR Haplotype Reference Database (YHRD) [8] and were uploaded to the database (submission accession numbers: YA005979-1 for English, YA005879-1 for Irish, YA006006-1 for Welsh, and YA006005-1 for Scottish). Haplogroups were determined using the Y-DNA Haplogroup Predictor - NevGen online tool [9]. Haplogroup diversity was calculated in Excel using the formula $n(1-\sum p_i^2)/(n-1)$, where p_i is the frequency of the i^{th} haplogroup, and n is the total number of samples [10].

Results

All samples produced unique Y-STR haplotypes. The NevGen prediction revealed a total of 46 haplogroups, with R1b M269 being the most common group across all four populations (Figure 1). The second most common haplogroup was I1 in the English and Scottish populations, while I2a2a M223 was more frequent in the Welsh and Irish samples. I2a2a M223 was also the third most common haplogroup in the English dataset. Overall, R1b M269 was observed in 73% of the samples, I1 in 8%, I2a2a M223 in 5%, followed by R1a M198 (3%) and E1b1b > V13 (2%). The remaining 41 haplogroups were found fewer than six times each across all populations (<0.6%).

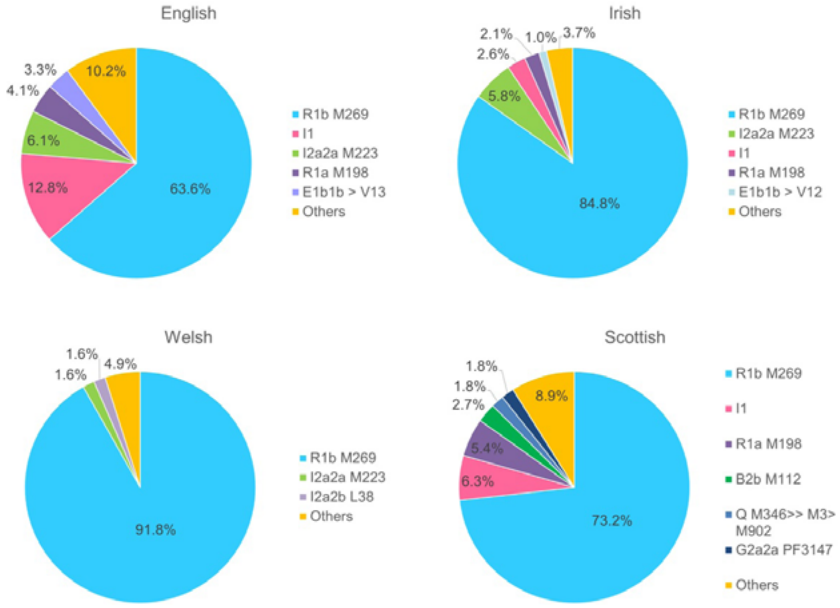


Figure 1: Pie charts representing the frequency of the main haplogroups identified among the English, Irish, Welsh, and Scottish population samples.

The Welsh population, with over 90% of individuals belonging to R1b M269, exhibited the lowest diversity (Table 1). The Irish dataset showed the second lowest haplogroup diversity, followed by the Scottish and English populations. In comparison to the others, the Scottish population displayed a higher number of haplogroups relative to the sample size tested.

Table 1: Sample size, number of observed haplogroups, and haplogroup diversity for each population tested.

Population	Sample Size	Observed Haplogroups	Haplogroup Diversity
English	461	34	0.57
Irish	191	12	0.28
Welsh	124	9	0.16
Scottish	112	16	0.46

Discussion

The results suggest the presence of genetic differences in the Y chromosome across the four populations of the British Isles tested. R1b M269, also referred to as R1b1a2 (R-M269), a subclade of haplogroup R1b, has previously been reported as the major lineage in Western Europe [11-12]. There is a gradient of increasing prevalence towards the west, from 12% in eastern Turkey to 85% in Ireland, with approximately 110 million European men carrying this haplogroup. The time to the most recent common ancestor (TMRCA) was also estimated, with the oldest value recorded in central Turkey (7,989 years [95% confidence interval (CI): 5,661–11,014]) and the youngest in Cornwall (5,460 years [3,764–7,777]) [12]. Some populations in Cornwall and Wales have also been observed to be fixed for the R1b haplogroup [13], which aligns with the findings of the present study.

The I1 haplogroup, also known as I-M253, has been identified as a native European haplogroup associated with Germanic migrations, with a higher prevalence in England compared to other regions of the British Isles [14]. This haplogroup is also found in 37-39% of Swedish [15], Norwegian [16], and Danish [17] males, making it a useful marker for "invaders", such as the Vikings or Anglo-Saxons. As observed in this study, previous research has shown a higher frequency of I1 in England compared to other parts of the British Isles [13].

Overall, a significant patrilineal population structure has been observed in areas influenced by Anglo-Saxon migration [13]. Similarly, genetic data suggests that the eastern part of the British Isles, particularly England, has experienced increased genetic influence due to multiple migrations and invasions, including those of the Anglo-Saxons [18]. Conversely, Ireland, Wales, and Scotland are believed to have retained a higher proportion of original Celtic DNA. The Welsh gene pool, in particular, is thought to be the closest to that of the first settlers who arrived in Britain after the last Ice Age [19]. As expected, this study shows greater Y haplogroup diversity in the English dataset compared to other areas of the British Isles, while the Welsh population exhibits the lowest diversity, followed by the Irish and Scottish populations.

While this study confirms previous findings and highlights the need for separate databases to perform statistical calculations related to Y-STRs, some limitations should be noted. Firstly, the small size of the datasets, particularly for the Irish, Welsh, and Scottish populations, may have affected the observations and skewed the percentages. Secondly, the accuracy of the haplogroup predictions may have been impacted by the prediction tool used. Inaccurate estimates have been

reported when using 19 Y-STRs to predict haplogroups with various prediction tools, including NevGen [5]. A combination of STRs and SNPs, with a higher number of markers, may yield more accurate haplogroup determinations and should be considered for future studies [4].

Conclusion

This study highlights the differences between Y haplogroups in the four populations of the British Isles, reaffirming the need to establish local Y-STR databases for the English, Irish, Welsh, and Scottish populations. While the distribution of haplogroups supports previous findings, further studies with more markers and larger sample sizes are required to continue populating Y-marker databases and to ensure the most accurate results. A population-specific Y-STR database would be invaluable in both legal cases (e.g., paternity or criminal cases) and historical studies tracing lineage and migration. This would significantly expand the application of Y chromosome testing, particularly in forensics, genealogy, and historical research.

Conflict of interest statement

All claims made in this article are solely those of the authors and do not necessarily reflect the views of their affiliated company or organisation.

References

[1] S. Gopinath, C. Zhong, V. Nguyen, et al., Developmental validation of the Yfiler® Plus PCR Amplification Kit: An enhanced Y-STR multiplex for casework and database applications, Forensic Sci. Int. Genet. 24 (2016) 164-175.

[2] D. Syndercombe Court, The Y chromosome and its use in forensic DNA analysis, Emerg. Top. Life Sci. 5 (2021) 427-441.

[3] M. Kayser, Forensic use of Y-chromosome DNA: a general overview, Hum. Genet. 136 (2017) 621–635.

[4] V.N. Kharkov, Y-Chromosome Markers in Population Genetics: Fundamental and Applied Results of Ethnogenomic Research, Russ. J. Genet. 57 (2021) 989-1001.

[5] B. Emmerova, E. Ehler, D. Comas, et al., Comparison of Y-chromosomal haplogroup predictors, Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) e145-e147.

[6] J.M. Butler, Chapter 9 – Y Chromosome DNA Testing. In: J.M. Butler, Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers, Second Edition, Elsevier Academic Press, New York, 2005, pp. 201-239.

[7] S. Iyavoo, S. Knights, M. Mavrommatis, et al., Implementation of Prep-n-Go™ Buffer for DNA extraction from buccal swabs. Forensic Sci. Int. Genet. Suppl. Ser. 7 (2019) 211-213.

[8] S. Willuweit, L. Roewer, The new Y Chromosome Haplotype Reference Database, Forensic Sci. Int. Genet. 15 (2015) 43-48.

[9] M. Cetkovic Gentula, A. Nevski, Y-DNA Haplogroup Predictor—NevGen. 2015, Available at: https://www.nevgen.org/ (Accessed 28 September 2024).

[10] M. Nei, F. Tajima, DNA polymorphism detectable by restriction endonucleases, Genet. 97 (1981) 145-163.

[11] International Society of Genetic Genealogy. Haplogroup R1b (Y-DNA) - ISOGG Wiki, 2021. Available at: https://isogg.org/wiki/Haplogroup_R1b_(Y-DNA) (Accessed 28 September 2024).

[12] P. Balaresque, G.R. Bowden, S.M. Adams, et al., A Predominantly Neolithic Origin for European Paternal Lineages, PLOS Biol. 8 (2010) e1000285.

[13] T.I. Huszar, W.F. Bodmer, K. Hutnik, et al., Sequencing of autosomal, mitochondrial and Y-chromosomal forensic markers in the People of the British Isles cohort detects population structure dominated by patrilineages, Forensic Sci. Int. Genet. 59 (2022) 102725.

[14] G. Broich, M. Fabbri, F.M. Avato, Specific European Y-Chromosome Haplotype I and its subclasses: migrations and modern prevalence, IXth Colloque of Genealogy, Madrid, 21-24 October 2015.

[15] T. Lappalainen, U. Hannelius, E. Salmela, et al., Population Structure in Contemporary Sweden—A Y-Chromosomal and Mitochondrial DNA Analysis, Ann. Hum. Genet. 73 (2009) 61–73.

[16] B.M. Dupuy, M. Stenersen, T.T. Lu, et al., Geographical heterogeneity of Y-chromosomal lineages in Norway, Forensic Sci. Int. 164 (2006) 10–19.

[17] J.J. Sanchez, C. Børsting, A. Hernandez, et al., Y chromosome SNP haplogroups in Danes, Greenlanders and Somalis, Int. Congr. Ser. 1261 (2004) 347–349.

[18] J. Perry, S. Pedroza Matute, S. Cummings, et al., English and Irish population comparison using STR markers: Insights into genetic disparities and historical influences, Forensic Sci. Int. Rep. 8 (2023) 100334.

[19] S. Leslie, B. Winney, G. Hellenthal, et al., The fine-scale genetic structure of the British population, Nature 519 (2015) 309–314.

30th INTERNATIONAL SOCIETY FOR FORENSIC GENETICS CONGRESS
Universidade de Santiago de Compostela, 2025, pp. 305-311
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Forensic and population genetic analysis of the X-STR loci in the ForenSeq™ Signature Prep Kit in the Ghanaian population

Pet-Paul Wepeba*[1,2], William Goodwin[3]

[1]*School of Life Sciences Anglia Ruskin University, Cambridge*
[2]*Africa Centre for Human ID and Genomic Medicine, GH*
[3]*School of Law and Policing, University of Central Lancashire. Preston*

WHGoodwin@uclan.ac.uk
Corresponding Author: Pet-Paul Wepeba, Pet-Paul.Wepeba@aru.ac.uk*

Abstract
Forensic genetic studies of the X-chromosome are limited for the populations of Sub-Saharan Africa. In Ghana, there is currently little forensic genetic data available on the X-chromosome for forensic applications and population characterisation. The ForenSeq™ Signature Prep kit was used in this study to obtain length and sequence-based data from 100 unrelated male samples from Ghana's four main ethnic groups: the Akan, the Ewe, the Mole-Dagomba, and the Ga-Adangbe. The DXS10074 showed the highest per locus average DoC of 1608.62. The total number of alleles generated was 179, 211, and 217 for length, repeat, and flanking region sequence datasets, respectively, with allele ranges between 7 and 40.2. The DXS10074 and DXS10135 loci were the most informative markers, while DXS7423 was the least. For length-based X-STRs, the observed number of alleles per locus ranged from 4 for the DXS8378 and DXS7423 loci to 19 for DXS10135. The repeat region sequenced data produced alleles ranging from 4 for the markers DXS8378 and DXS7423 to 44 for DXS10135. The increase in the number of alleles and heterozygosity impacted and enhanced the power of discrimination from 70% to 90%. The pairwise analysis revealed that the Mole-Dagomba, Ewe, and Ga-Adangbe subpopulations, when compared, all had low FST values ranging from 0.02 to 0.03 for both length and sequence-based data; all four sub-populations clustered in a PCA. Lastly, the study demonstrated that X-STRs would be effective and suitable for complex kinship analysis, complementing the autosomal STR profiling for human identification, due to their power of discrimination in the Ghanaian population.

Keywords
Ghana, genetics, sequence, X-chromosome, STRs.

1. Introduction

The X-chromosome short tandem repeats (STR) makers are a tool that can be used in complex kinship and paternity analysis for females when the alleged father(s) are absent and in incest, or if autosomal STR testing fails to yield sufficiently informative results [1]. Using between six and twelve X-chromosome STR markers along with autosomal STR results can improve the discrimination index [2,3]. This is especially useful for ruling out a false father when there is a trio in which the child is a female or in maternity testing [4]. The challenge with X-chromosome analysis is that, because the X-chromosome map covers 240 cM, the selection of more than five unlinked markers may not be possible. This makes it unsound to use the product rule in the multi-locus likelihood ratio calculations. However, the X-chromosome markers can still be useful in complex kinship testing [5]. This study aimed to sequence 100 unrelated male samples from the main ethnic groups in Ghana using the ForenSeq™ signature prep kit.

2. Materials and methods

2.1. Sample collection, purification and quantification

A 100 unrelated male DNA samples were collected with informed consent from the four main ethnic groups of Ghana; Akan, Mole-Dagomba, Ewe and Ga-Adangbe (25 per subpopulation). Samples were purified using the Qiagen puregene DNA extraction kit and quantified using the Quantifiler Trio kit on the QuantStudio 5 Real Time PCR thermal cycler, following the manufacturer's recommendations.

2.2. Amplification, sequencing and data analysis

The primer mix A (DPMA) from the ForenSeq™ DNA Signature Prep Kit was used to amplify and tag the target genome for 100 samples. This was done in six steps, as recommended by the manufacturer. However, the normalised pooled library volume added to the Human Sequencing Control (HSC) was increased from 7 µl to 12 µl, as used by Devesse *et al* [6]. We denatured and diluted the pooled libraries as per the manufacturer's guide, loaded them into a microchip flow cell, and then sequenced them on the MiSeq FGx sequencer to generate data from Primer set A (DPMA; 153 loci) comprises 58 STRs, 94 identity-informative SNPs, and amelogenin in a single PCR multiplex reaction [7].

The ForenSeq™ Universal Analysis Software (UAS) was used to analyse samples on the MiSeq™ FGx, applying default settings for the analytical and interpretation thresholds. We used Microsoft Excel to retrieve alleles and sequences from the UAS data and filter out novel variants. We searched for novel alleles in the 1000 genome browser using the GRCh37 genome assembly, and the ISFG nomenclature system [8,9]. We calculated the depth of coverage (DoC) by summing all effective locus reads. We used the Arlequin v3.5.2 software [10] to check linkage disequilibrium between all 121 autosomal markers in the ForenSeq™ DNA Signature Prep. We analysed the forensic parameters of the X-STRs using an online tool (http://www.chrx-str.org/) and the STRAF software for population parameters [11].

3. Results

The test yielded a cluster density of 1148 K/mm², with 91.78% passing the Illumina chastity filter. The run-metric indicators were within the recommended range of 400-1650 K/mm², with ≥ 80%, ≤ 0.25%, and ≤ 0.15%. Read 1, index 1, and index 2 all passed the quality filter, as shown in the run metric report.

The per locus coverage graph in Figure 1 showed the DXS10074 and the DXS10103 loci with the maximum and minimum coverage. The read coverage of the DXS10103 locus was however, higher than the minimum analytical, and interpretation thresholds and can thus be used.

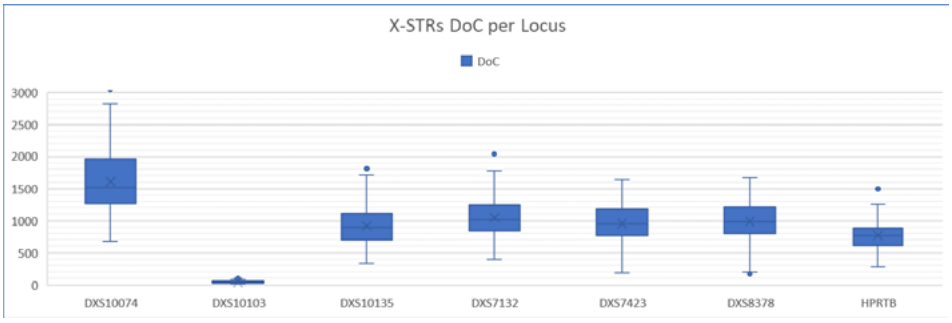


Figure 1. Depth of coverage for the 7 X-STR loci tested for the 100 samples in the four subpopulations combined in this study. The DXS10103 had the fewest number of read and DXS10074 the highest read count. Each box plot represents the depth of coverage for each locus for each of the 100 samples.

We observed a total of 179, 211, and 217 alleles for the length-based, repeat region, and repeat and flanking region sequences of the X-STRs, respectively. The X-STR allele range was between 7 and 40.2. The highest allele frequencies observed for the Akan and Ewe were 0.64 and 0.60, respectively, for DXS7423. The dominant allele 19 in the Ga-Adangbe produced an allele frequency of 0.60 for the DXS10103 locus, and the Mole-Dagomba produced a frequency of 0.52 for the dominant allele 14 in the DXS7423 locus.

We calculated the power of discrimination (PoD) for both the homozygous XX and the XY hemizygous scenarios. We compared the PoD of the four subpopulations using the length and sequence-based polymorphic data. The DXS10135 was the locus with the highest PoD for both length and sequence-based data in the male population. For length-based X-STRs, the observed number of alleles per locus ranged from 4 for the DXS8378 and DXS7423 loci to 18 for DXS10135. The repeat region sequenced data produced alleles ranging from 4 for the markers DXS8378 and DXS7423 to 44 for DXS10135. The loci DXS7423 and DXS10135 had an allele range of 4 to 45 for the repeat and flanking region sequenced data.

Sequence-based data from the X-STRs led to an increase of 26 alleles in the locus DXS10135, which improved the PoD in the XY hemizygous group from 70% to 90%. The pairwise F_{ST} values on X-STRs ranged from 0.02 to 0.03, as showed in Table 1. The increase in sequence-based F_{ST} values may be due to rare sequence variants rather than the small sample size. Principal Component The analysis of the four sub-populations using the X-STRs length, repeat, and flanking region sequence-based data revealed that all four subpopulations had a closed genetic structure, with no significant differentiation between them.

Table 1. Pairwise F_{ST} generated using the STRAF software (http://cmpg.unibe.ch/shiny/STRAF/) for the length and sequenced data of the X-STRs for genetic differentiation of the Akan, Ewe, Ga-Adangbe and Mole-Dagomba. Using the X-STRs length based and length and sequence-based data as indicated in brackets in the table, both produced low F_{ST} values ranging from 0.02 to 0.03.

	Akan	Ewe	Ga-Adangbe	Mole-Dagomba
Akan	*			
Ewe	0.02 (**0.03**)	*		
Ga-Adangbe	0.03 (**0.03**)	0.02 (**0.02**)	*	
Mole-Dagomba	0.03 (**0.03**)	0.02 (**0.02**)	0.02 (**0.02**)	*

4. Discussion

The number of observed alleles per locus for the length-based dataset ranged from 4 to 9. The sequence-based data for the observed alleles repeat region ranged between 4 and 21. The repeat and flanking region sequence-based data observed allele range from 4 to 28. Other studies using the ForenSeq™ DNA Signature Prep Kit have reported similar results [12,13].

The DXS10074 and DXS8378 had a >66% increase in the number of alleles, and the DXS10135 locus had an increase of 136.8%. These loci significantly influenced the discrimination power. The DXS10074 and DXS10135 loci were the most informative markers, while DXS7423 was the least informative using the X-STRs. A study that looked at the datasets by locus showed that the observed alleles and heterozygosity of the two multi-loci affected and improved the ability to differentiate the two X-STRs [12]. All four subpopulations did not demonstrate significant differentiation with their pairwise F_{ST} values. The F_{ST} values observed by the X-STRs ranged between 0.02 and 0.03. The Ewe showed the highest X-STR F_{ST} of 0.02 against the Ga-Adangbe, and Mole-Dagomba in both the length and sequence-based data, but 0.03 against the Akan using the sequenced-data. F_{ST} value of 0.01 for X-STRs have been observed in a Spanish population using the ForenSeq™ DNA Signature Prep Kit [12].

Principal component analysis of the X-STRs revealed that all four sub-populations were close in genetic structure. The Mole-Dagomba and the Ewe practice patrilineal and patrilocal systems, which could explain the mild genetic structure in part. Generally, the F_{ST} values between the four studied sub-populations were low. DNA polymorphisms in the lineage markers are critical in population genetics. Ballantyne [14] use information from the analysis of genetic markers on the X-chromosome to infer the evolutionary and genetic genealogy of individuals. In particular, the X-chromosome is efficient in population genetics due to its low mutation and recombination rate and relatively fast genetic drift. The X-chromosome can thus provide a clearer population structure than autosomes [15]. In this study, the per locus analysis revealed that the observed alleles and heterozygosity enhanced discrimination power. The lineage markers complement the autosomal markers in forensic applications. The X-chromosome is useful in complex paternity cases [16].

5. Conclusion

The study collected MPS lineage marker data from 100 Ghanaian ethnic groups using Primer Mix A and MiSeq FGx Genetic Sequencer. The data showed no significant differentiation between sub-populations, and sequence-based data improved the PoD.

6. Acknowledgments

This project was funded by the Ghana Education Trust Fund

7. Conflict of interest statement

None

8. References

[1] U. Toscanini, G. Berardi, A. Gómez, E. Raimondi, X-STRs analysis in paternity testing when the alleged father is related to the biological father, Foren. Sci. Int. Genet Suppl. Ser. 2(1) (2009) 234-5.

[2] J. Aquino, C. Peixe, D. Silva, C. Tavares, EF de Carvalho, X-chromosome STR hexaplex as a powerful tool in deficiency paternity cases, Foren. Sci. Int. Genet Suppl. Ser. 2(1) (2009) 45-6.

[3] A.O. Tillmar, C. Phillips, Evaluation of the impact of genetic linkage in forensic identity and relationship testing for expanded DNA marker sets, Foren. Sci. Int. Genet. 26 (2017) 58-65.

[4] E.M. Rodrigues, F.P. das Neves Leite, M.H. Hutz, T.D. Palha, Â.K. dos Santos, S.E. dos Santos, A multiplex PCR for 11 X chromosome STR markers and population data from a Brazilian Amazon Region, Foren. Sci. Int. Genet. 2(2) (2008) 154-8.

[5] S. Presciuttini, C. Toni, M. Alu, A. Asmundo, L. Baldassarri, A. Barbaro, L. Caenazzo, E. Carnevali, N. Cerri, E. D'Aloia, C. Di Nunzio, X-chromosome in Italy: A database of 29 STR markers, Foren. Sci. Int. Genet Suppl. Ser. 3(1) (2011) e37-8.

[6] L. Devesse, D. Ballard, L., Davenport, I. Riethorst, G. Mason-Buck, D.S. Court, Concordance of the ForenSeq™ system and characterisation of sequence-specific

autosomal STR alleles across two major population groups, Foren. Sci. Int., Genet. 34 (2018) 57-61.

[7] A.C. Jäger, M.L. Alvarez, C.P. Davis, E. Guzmán, Y. Han, L. Way, P. Walichiewicz, D. Silva, N. Pham, G. Caves, J. Bruand, Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories, Foren. Sci. Int. Genet. 28 (2017) 52-70.

[8] 1000 Genomes Project Consortium [online]: rs740910 (SNP)-Population Genetics-Homo sapiensEnsemblgenomebrowser100.Available at:(http://www.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=17:5802803-5803803;v=rs740910;vdb=variation;vf=87302655) [Accessed: 5 March 2020].

[9] A.O. Tillmar, D. Kling, J.M Butler, W. Parson, M. Prinz, P.M. Schneider, T. Egeland, L. Gusmão, DNA Commission of the International Society for Forensic Genetics (ISFG): Guidelines on the use of X-STRs in kinship analysis, Foren. Sci. Int. Genet. 29 (2017) 269-75.

[10] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5, A new series of programs to perform population genetics analyses under Linux and Windows, Mol. Eco. resourc. 10(3) (2010) 564-7.

[11] A. Gouy, M. Zieger, STRAF, A convenient online tool for STR data evaluation in forensic genetics, Foren. Sci. Int. Genet. 30 (2017) 148-51.

[12] F. Casals, R. Anglada, N. Bonet, R. Rasal, K. J. van der Gaag, J. Hoogenboom, N. Solé-Morata, D. Comas, F. Calafell, Length and repeat-sequence variation in 58 STRs and 94 SNPs in two Spanish populations, Foren. Sci. Int. Genet. 30 (2017) 66-70.

[13] A. Delest, D. Godfrin, Y. Chantrel, A. Ulus, J. Vannier, M. Faivre, C. Hollard, F.X. Laurent, Sequenced-based French population data from 169 unrelated individuals with Verogen's ForenSeq DNA signature prep kit, Foren. Sci. Int. Genet. 47 (2020) 102304.

[14] K.N. Ballantyne, A. Ralf, R. Aboukhalid, N.M. Achakzai, M.J. Anjos, Q. Ayub, J. Balažic, J. Ballantyne, D.J. Ballard, B. Berger, C. Bobillo, Toward male individualization with rapidly mutating Y-chromosomal short tandem repeats, Hum mut. 35(8) (2014) 1021-32.

[15] Y. Li, X. Li, W. Chen, Y. Fan, M. Xie, J. Wu, Allele and haplotype frequencies of 19 X-STRs in the Kyrgyz and Han populations from Kizilsu Prefecture, Foren. Sci. Int. Genet. 40 (2019) e259-61.

[16] T.M. Diegoli, Forensic typing of short tandem repeat markers on the X and Y chromosomes, Foren. Sci. Int. Genet. 18 (2015) 140-51.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 313-317
DOI: https://dx.doi.org/10.15304/cc.2025.1869

A case study of a male inferred to be of Japanese origin based on Y-STRs haplotypes

Toshimichi Yamamoto[1*], Takashi Yoshimoto[1], Yujin Natori[1], Akira Ishii[1]

[1] *Department of Legal Medicine and Bioethics, Nagoya University, Nagoya, Japan.*

** Department of Legal Medicine and Bioethics, Graduate School of Medicine, Nagoya University, 65 Tsurumai-cho, Showa-ku, Nagoya, 466-8550 Japan,*

Tel.: +81-52-744-2119; Fax: +81-52-744-2121, E-mail; yamachan@med.nagoya-u.ac.jp

Abstract

Even today, some 80 years after World War II, children born to Japanese men, such as Japanese businessmen, and Filipino women before and after the war (Filipino residual orphans) are still alive and reaching old age. In fact, according to a survey conducted by the Ministry of Foreign Affairs (FY 2018), there are approximately 1,000 survivors whose Japanese nationality has not been restored and who remain stateless, regardless of whether their father's identity has been determined or not. One of them (Mr. A), with the cooperation of a certain NPO, filed a lawsuit with the court, but the case was dismissed because there was no particular documentary proof such as family register. A company was asked to test the Y-STR haplotype (y-ht) and the Y haplogroup (y-hg), but the y-ht did not match any possible person in the investigation. We appealed to the high court without any new and glaring evidence, but the appeal was dismissed. So, I was consulted to see if there was any way to analyze the situation. His y-ht at 15 out of 17 Y-STRs of Mr. A's Yfiler tested by a certain company was used for population genetical analyses. As controls for Japanese and Filipinos, data of 1280 and 790 y-hts, respectively, were analyzed by excluding y-hts including null and duplicated alleles from the literature and web databases. A total of 2071 y-hts of data from these 15 loci were used for principal component analysis and network analysis. Comparing the y-hts of Mr. A with those of Japanese and Filipinos, the y-hts of Mr. A were characteristic of Japanese, and the Y-chromosome inherited by Mr. A was considered to be derived from Japanese, not Filipinos. The response letter also added that, since the survivor in question was elderly, it would be possible to estimate the generation to some extent through genome-wide analysis, etc., as long as DNA samples were preserved in the prenatal data. This written response

was submitted to the High Court as new evidence and a retrial was requested, which was granted, and Mr. A was granted Japanese nationality.

Keywords
Filipino residual orphans, Japanese nationality, Y-STR haplotype, Population genetics.

Introduction

Even today, about 80 years after World War II, children born to Japanese men, such as Japanese businessmen, and Filipino women during the prewar and postwar periods are surviving and reaching old age. In fact, a survey by the Ministry of Foreign Affairs (FY 2018) shows that there are approximately 1,000 survivors whose Japanese nationality has not been restored and who remain stateless, regardless of whether the father's identity has been determined not, because there was no particular documentary proof such as Japanese nationality by the father's name unknown.

One of them (Mr. A), with the cooperation of a certain NPO, filed a lawsuit with the court, but the case was dismissed because there was no obvious documentary proof. A company was asked to test the Y-STR haplotype (y-ht) and the Y haplogroup (y-hg), but the y-ht did not match any possible person in the investigation and he could not be estimated as Japanese origin from his y-hg. Nevertheless, they appealed to the high court without any new and glaring evidence, but the appeal was dismissed.

Subsequently, the non-profit organization was looking for a researcher who might be able to extrapolate the Japanese from that y-ht. I was asked by one of the professors if there was any possibility to solve it and attempted a preliminary analysis. The results showed that it could be presumed to be of Japanese origin, and the lawyer for that NPO asked me to prepare a formal report.

Material and Methods

Material (Data)

2.2.1 Mr. A's Y chromosome STRs haplotype

Y-STRs at 17 loci using Yfiler and Y haplogroup (O-P45)

2.1.2 Y-STRs data at 17 loci in general Japanese males [1]

(1) 25 y-hts using Yfiler plus from 1299 Japanese males in Hokkaido, Honshu, Shikoku, Kyushu and Okinawa.

(2) 1218 out of 1299 y-hts associated with the data of Y-hgs (C-M130*, C-M131, D-M174*, D-M57*, D-M125, N-M231, O-M175*, O-M122*, O-M134 and Q-P36.2).

(3) A total of 1280 data were used by excluding null alleles at DYS448 and duplicate alleles at DYS389.

2.1.3 Y-STRs data for 17 loci in the general Filipino male population

(1) Y-STRs typing results of 169 Filipino men (Philippines [Filipino]) at 23 loci selected from 19,630 men from 129 populations in 51 countries worldwide using the PowerPlex Y23 System [2].

(2) Y-STRs at 23 loci in 629 Filipino males obtained personally from the registrant (National Police and Crime Laboratory) registered at YHRD (https://yhrd.org/) (Submission Accession Number: YA003892-1)

(3) A total of 790 names were used by excluding null allele (DYS448), etc.

Population genetical analysis

2.2.1 PCA (Principal Component analysis) using STRAF1.0.5 [3]

15 out of 17 Y-STR used without DYS385a/b and with DYS389 (II-I) and a triangle plot was performed using JMP 14.0 software.

2.2.2 Network analysis using Network 10.2.0.0 and Network publisher 2.1.2.5 [4]

15 out of 17 Y-STR used without DYS385a/b and with DYS389(II-I) and) The networks were constructed with y-hgs and countries by Median-joining method.

Results and discussion

The results of the PCA and the triangular plots using the three PCs showed that the groups were broadly classified into two groups: Japanese-only and mixed groups from both countries; although Mr. A was in the former group, one Filipino plot was

present in close proximity. We therefore examined the haplotypes of that plot and found that it differed from Mr. A's by a total of 15 alleles.

Subsequently, the network analysis with country information (Fig. 1A) and y-hgs information (Fig. 1B) revealed a group consisting almost exclusively of Japanese with hg-O-M175*, and Mr. A's node was in this group. Therefore, we zoomed in on the orange dashed line in Fig. 1A (Fig1.C) and observed two Japanese nodes (JPN0558 and 0020) and one Filipino node (Filn059) near Mr. A's node. We examined their haplotypes and found that they differed from Mr. A by a total of 4, 6, and 12 alleles, respectively. From these results, we considered that Mr. A was derived from a Japanese male and prepared a report.

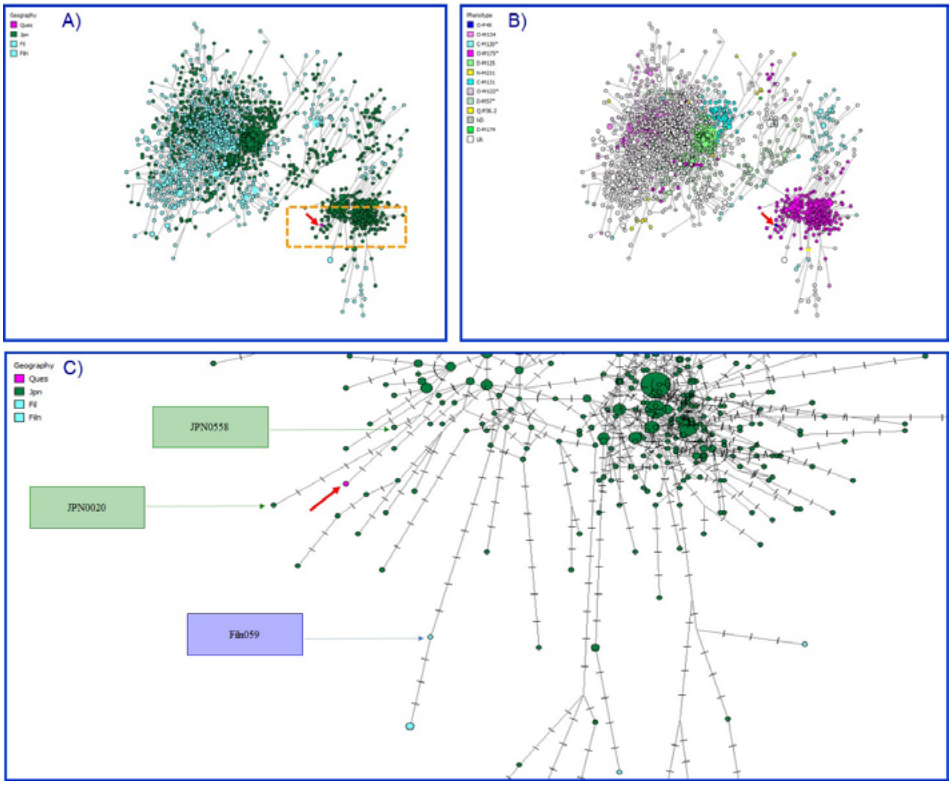


Fig. 1. The result of Network analysis. The red arrows indicate the plot of Mr. A. A network with country. B; A network with haplotypes. C; Enlarged view of the area enclosed by the orange dashed line.

I also stated as an opinion in the report that a genome-wide SNP analysis would have yielded clearer results, but due to financial and ethical issues, the project has not progressed as far as I would have liked. However, a case like Mr. A's was dismissed by two of his siblings, who are over 80 years old, and a few months ago, the same method was used to analyze the case, and the results were considered to be of Japanese origin. The case is still pending in court.

Conclusion

Based on this report, a request for a retrial was filed, and Mr. A was admitted as a Japanese citizen. Unfortunately, however, Mr. A died about a year after acquiring Japanese nationality.

Acknowledgments

We sincerely thank Dr. Sheila Angustia for sending us these valuable data about Y-STR haplotypes of Filipino males.

Conflict of interest statement

No conflict of interest.

References

[1] H. Watahik, K. Fujii, T. Fukagawa, Y. Mita, T. Kitayama,N., Mizuno, Polymorphisms and microvariant sequences in the Japanese population for 25 Y-STR markers and their relationships to Y-chromosome haplogroups. FSI: Genet 41 (2019) e1–e7.

[2] J. Purps, S. Siegert, S. Willuweit, M. Nagy et al. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci, FSI: Genet12 (2014) 12–23.

[3] STRAF1.0.5, https://straf.fr/

[4] Network analysis using Network 10.2.0.0 and Network publisher 2.1.2.5, https://www.fluxus-engineering.com/sharenet.htm.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 319-325
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Characterization of 94 Identity-Informative SNPs in a North-Eastern Italian population using ForenSeq™ DNA Signature Prep Kit

Chiara Saccardo[1*], Francesco Ausania[1], Giulia Soldati[1], Domenico De Leo[1], Stefania Turrina[1]

[1]*Department of Diagnostics and Public Health, Section of Forensic Medicine, Forensic Genetics Lab, University of Verona, Verona, Italy.*

[*]*Corresponding author at: Policlinico G.B. Rossi, P.le L.A. Rossi, 37134 – Verona (Italy), chiara.saccardo@univr.it, telephone number: +39 3497021566*

Abstract

Genotyping of identity-informative single nucleotide polymorphisms (iiSNPs) by massively parallel sequencing (MPS) provides information on the sequence variations that, characterizing the individual populations, may represent a valid support for the analysis of short tandem repeats (STRs) with capillary electrophoresis (CE) in some challenging forensic cases, such as complex kinship analyses and mixed trace deconvolution. Nonetheless, to date, population data on iiSNPs are relatively scarce, especially concerning the Italian population.

In this study, using the MPS technique, 93 subjects from North-East Italy were genotyped with the 94 iiSNPs included in the DNA Primer Mix A of the ForenSeq™ DNA Signature Prep kit (Verogen) to obtain allele frequencies and relevant forensic statistic parameters based on sequence variations revealed in the target and the flanking regions of the iiSNPs.

Keywords

Massively parallel sequencing (MPS), Identity informative single nucleotide polymorphisms (iiSNPs), Microhaplotypes (MHs), MiSeq™ FGx Forensic Genomics System.

Introduction

Massively parallel sequencing (MPS) undoubtedly represents an analytical resource to be exploited in the forensic genetics field that, due to its ability to ensure simultaneous analysis of the length and sequence of the DNA fragments and typing in

a single reaction of a considerable number of different genetic markers, allows to optimize the amount of DNA available for the analysis.

Sequence variations revealed by MPS concern both target and flanking regions of the marker, representing the primary source of allelic diversity. Concerning Single Nucleotide Polymorphisms (SNPs), detecting additional SNPs in close proximity to the target, i.e., microhaplotypes (MHs) [1], may change their original biallelic nature into multiallelic, affecting the number of effective alleles (A_e) and those forensic parameters, such as the power of discrimination (PD), that contribute to better characterize the individual populations [2-4].

For the above reasons, in this study, the focus was on the sequence-based analysis of 94 identity-informative SNPs (iiSNPs) included in the DNA Primer Mix A of the ForenSeq™ DNA Signature Prep Kit (Verogen) for which allele frequencies and forensic parameters were determined for the first time in a population of North-East Italy.

Materials and methods

Anonymized saliva samples were collected from 93 unrelated subjects native to North-East Italy after receiving their written informed consent and ethical approval from the University of Verona's Research Ethics Committee (protocol code: CARU-12/2020).

Genomic DNA was extracted using the QIAamp® DNA Mini Kit (Qiagen) and quantified using the Qubit® dsDNA HS Assay Kit on Qubit® 2.0 Fluorometer (Thermo Fisher Scientific).

The DNA Primer Mix A from the ForenSeq™ DNA Signature Prep Kit (FSSP) (Verogen), which contains primers to amplify 94 iiSNPs [5], was used for MPS library preparation. Sequencing was carried out on the MiSeq FGx™ System (Illumina) [6], and data analysis was performed by Universal Analysis Software (UAS) v1.2 (Verogen), applying its default parameters and thresholds. This study employed a minimum locus threshold of 30 reads to guarantee that the unique sequences detected were non-fictitious.

Considering that version 1.2 of UAS is unable to generate the "Flanking Region Report", precluding the possibility of evaluating the presence of any sequence variations in the flanking regions of the SNPs, these were exclusively deduced based on the analysis of UAS's FASTQ files with the open-source bioinformatic tool STRait Razor v3 (SRv3) [7].

The comprehensive allelic nomenclature was assigned directly by SRv3, except for cases where it was attributed manually due to the detection of sequence variations not included in SRv3's string-matching database. In this instance, the rare or not described mutations were identified by aligning the sequences with the UCSC Genome Browser's human reference sequence (GRCh38).

Allele (haplotype) frequencies, effective number of alleles (A_e), heterozygosity (H_e), and relevant forensic statistical parameters, including the combined match probability (CMP) and combined power of discrimination (CPD), were calculated to assess how the presence of MHs affects the genetic variability of the iiSNPs.

Results and Discussion

Compared to the 8,742 expected call genotypes for the 93 subjects typed with the 94 iiSNPs of the FSSP kit, 8,571 genotypes were observed, corresponding to an overall call rate of 98%. The defection of 171 genotypes, due to a locus depth of coverage below the minimum threshold used, is mainly manifested as no-called genotypes or allelic dropout of a sister allele events in heterozygous genotypes that were incorrectly called homozygotes. In particular, low sequencing coverages were detected for rs1031825 and rs1736442, which were rightly genotyped in 66 and 60 subjects, respectively.

No discrepancies in target iiSNP genotype calls were detected when the number of reads for each iiSNP met the detection thresholds fixed for UAS and SRv3.

Allele frequencies and forensic parameters calculated for iiSNPs, when considered as biallelic markers, are reported in Table 1. An average heterozygosity of 0.4593, close to the maximum possible value of 0.5, was found, with CMP and CPD values of 5.50857E-38 and 1-5.50857E-38, respectively, in accordance with those provided by previous studies [2-4].

The detection of additional SNPs associated in close physical proximity to the target iiSNP, inferred by processing FASTQ files through SRv3, permitted the recognition, based on the minor allele (haplotype) frequency (MAF), of two SNP categories: microhaplotypes (MHs) and minor microhaplotypes (mMHs) [2]. For the 94 iiSNPs genotyped, SRv3 was able to perform haplotype calling of 19 MHs and 5 mMHs. However, for further 8 mMHs, SRv3 failed to recognize SNPs in the flanking regions, tagging them with NA, also considering that 6 of these have never been reported in the literature (reported in bold in Table 1); therefore, their nomenclature was manually assigned.

Table 1. Microhaplotypes (MHs) and minor microhaplotypes (mMHs) detected for 32 of the 94 iiSNPs investigated in a North-Eastern Italian Population

iiSNP	Haplotype category	Comprehensive Nomenclature	Haplotype frequency
rs1015250	MH	rs1015250 [CE C]-GRCh38-Chr9:1823726-1823793 rs6475200-G; rs1015250-C	0,129
rs10776839	MH	rs10776839 [CE G]-GRCh38-Chr9:134525445-134525507 rs7037930-G; rs10776839-G	0,269
		rs10776839 [CE T]-GRCh38-Chr9:134525445-134525507 rs7037930-G; rs10776839-T	0,473
rs1109037	MH	rs1109037 [CE G]-GRCh38-Chr2:9945582-9945659 rs1109037-G; rs1109038-A	0,161
		rs1109037 [CE A]-GRCh38-Chr2:9945582-9945659 rs1109037-A; rs1109038-A	0,339
rs12997453	mMH	rs12997453 [CE A]-GRCh38-Chr2:181548493-181548547 rs72883670-T; rs12997453-A	0,098
		[NA] rs12997453 [CE A]-GRCh38-Chr2:181548493-181548547 rs12997453-A	0,005
rs1336071	mMH	rs1336071 [CE C]-GRCh38-Chr6:93827491-93827549 93827520.1T; rs1336071-C	0,006
rs1463729	mMH	rs1463729 [CE C]-GRCh38-Chr9:124119138-124119195 rs1463729-C; rs553955089-C	0,005
rs2056277	mMH	rs2056277 [CE C]-GRCh38-Chr8:138386821-138386877 rs149318691-T; rs2056277-C	0,005
rs221956	mMH	**[NA] rs221956 [CE C]-GRCh38-Chr21:42186845-42186896 rs221956-C; 42186894>G**	0,005
rs2269355	mMH	**[NA] rs2269355 [CE C]-GRCh38-Chr12:6836737-6836759 rs2269355-C; rs773954858**	0,005
rs2399332	MH	rs2399332 [CE G]-GRCh38-Chr3:110582177-110582286 rs2399334-T; rs2399333-T; rs2399332-G	0,604
		rs2399332 [CE G]-GRCh38-Chr3:110582177-110582286 rs2399333-T; rs2399332-G	0,027
rs279844	MH	rs279844 [CE T]-GRCh38-Chr4:46327592-46327708 rs279844-T; rs279845-A	0,423
rs2830795	MH	rs2830795 [CE A]-GRCh38-Chr21:27235795-27235862 rs12626695-C; rs2830795-A	0,038
		rs2830795 [CE G]-GRCh38-Chr21:27235795-27235862 rs12626695-C; rs2830795-G	0,011
rs2920816	MH	rs2920816 [CE G]-GRCh38-Chr12:40469199-40469309 rs142684512-G; rs2920816-G	0,056
rs321198	mMH	rs321198 [CE T]-GRCh38-Chr7:137344994-137345108 rs321198-T; rs186446125-C	0,005
rs354439	mMH	[NA] rs354439 [CE T]-GRCh38-Chr13:106285996-106286115 rs354439-T; rs574448517-C	0,005
rs430046	MH	rs430046 [CE T]-GRCh38-Chr16:77983107-77983174 rs409820-A; rs430044-T; rs430046-T	0,435
rs445251	MH	rs445251 [CE C]-GRCh38-Chr20:15144244-15144311 rs369438-C; rs445251-C	0,548
rs4530059	MH	rs4530059 [CE A]-GRCh38-Chr14:104302784-104302911 rs4530059-A; rs4450333-T	0,403
rs4606077	MH	rs4606077 [CE C]-GRCh38-Chr8:143574562-143574669 rs4606077-C; rs1869434-A	0,775
		rs4606077 [CE T]-GRCh38-Chr8:143574562-143574669 rs4606077-T; rs58774517-T	0,066
		rs4606077 [CE C]-GRCh38-Chr8:143574562-143574669 143574562-G; rs4606077-C; rs1869434-A	0,016
rs6811238	mMH	**[NA] rs6811238 [CE G]-GRCh38-Chr4:168742417-168742480 rs6811238-G; rs1170846661**	0,005
rs6955448	MH	rs6955448 [CE T]-GRCh38-Chr7:4270677-4270749 rs6950322-A; rs6955448-T	0,183
rs7041158	MH	rs7041158 [CE T]-GRCh38-Chr9:27985930-27986001 rs7041158-T; rs117776421-G	0,011
rs722098	mMH	**[NA] rs722098 [CE A]-GRCh38-Chr21:15313267-15313315 rs722098-A; rs1036208975**	0,005
rs727811	MH	rs727811 [CE G]-GRCh38-Chr6:164624247-164624311 rs1390470-T; rs727811-G	0,478
		rs727811 [CE T]-GRCh38-Chr6:164624247-164624311 rs1390470-T; rs727811-T	0,522
rs740910	MH	rs740910 [CE A]-GRCh38-Chr17:5803260-5803323 rs60810599-G; rs740910-A	0,082
rs8078417	mMH	rs8078417 [CE T]-GRCh38-Chr17:82503992-82504093 82503995-C; rs8078417-T	0,005
		[NA] rs8078417 [CE C]-GRCh38-Chr17:82503992-82504093 rs8078417-C; rs1036208975	0,005
rs826472	mMH	rs826472 [CE C]-GRCh38-Chr10:2364342-2364443 rs138540224-C; rs826472-C	0,005

rs876724	MH	rs876724 [CE C]-GRCh38-Chr2:114970-115036 rs876724-C; rs300773-T	0,452
		rs876724 [CE T]-GRCh38-Chr2:114970-115036 rs876724-T; rs300773-T	0,349
rs907100	MH	rs907100 [CE C]-GRCh38-Chr2:238654924-238654989 rs907100-C; rs11689319-A	0,31
rs987640	MH	rs987640 [CE A]-GRCh38-Chr22:33163486-33163560 rs17793354-C; rs987640-A	0,075
		[NA] rs9905977 [CE G]-GRCh38-Chr17:3016058-3016176 rs9905977-G; rs187737298	0,005
rs9905977	mMH	rs9905977 [CE G]-GRCh38-Chr17:3016058-3016176 rs9905977-G; rs73298992-T	0,151
		rs9905977 [CE G]-GRCh38-Chr17:3016058-3016176 rs9905977-G; rs28582109-A	0,005
rs9951171	MH	rs9951171 [CE G]-GRCh38-Chr18:9749816-9749890 rs145524126-C; rs9951171-G	0,022

Overall, for the 32 iiSNPs presenting MHs and mMHs were detected 44 unique haplotypes spread in 966 out of 8,571 genotypes called, thus converting more than a third of the iiSNPs genotyped from biallelic to multiallelic and significantly affecting the CMP and CPD, that rise from 4.61964E-13 to 2.10942E-15, and from 0.9999999999995380 to 0.9999999999999980, respectively.

As an expression of genetic variability, it was assessed whether the detection of "new alleles" due to MHs and mMHs may influence A_e, H_e, and PD. The evaluation was performed on the 32 iiSNPs with additional SNPs in the flanking regions, assuming the markers as biallelic and multiallelic.

It was possible to observe that the transition from biallelic to multiallelic markers involved 81.25% of 32 iiSNPs, for which the number of effective alleles increased influencing (based on the type and frequency of the allele (haplotype) identified in the population) the He and PD values.

For instance, at rs727811, two MHs were detected due to the presence of rs1390470 in the flanking region of both wild type alleles G and T, but these latter were never found in the investigated population, thus A_e (equal to 2.00) remained unchanged, not affecting H_e and PD values. Regarding rs1336071, three alleles were revealed: two wild types (C and T) and a further allele due to mMH (SNP in position 93827520.1T). In this case, the number of effective alleles exceeded 2.00, but since the mMH had a frequency of 0.006, the increments in H_e and PD values were very low. Therefore, detecting a new but rare allele (haplotype) with a frequency < 0.01 does not seem to affect the H_e and PD values; conversely, significantly higher H_e and PD values were observed only for A_e greater than 2.50 (Table 2).

Table 2. Relevant forensic parameters for 32 of the 94 iiSNPs investigated (A_n: number of alleles; A_e: number of effective alleles; H_e: Heterozygosity; MP: Match Probability; PD: Power of Discrimination).

iiSNP	Target iiSNPs					MHs and mMHs				
	A_n	A_e	H_e	MP	PD	A_n	A_e	H_e	MP	PD
rs1015250	2	1,50	0,331	0,505	0,495	3	1,540	0,352	0,463	0,537
rs10776839	2	1,99	0,499	0,364	0,636	3	2,760	0,637	0,199	0,801
rs1109037	2	1,94	0,483	0,371	0,629	4	3,760	0,734	0,120	0,880
rs12997453	2	1,88	0,469	0,392	0,609	4	2,110	0,526	0,305	0,695
rs1336071	2	2,00	0,499	0,394	0,606	3	2,020	0,505	0,383	0,617
rs1463729	2	1,96	0,489	0,356	0,644	3	1,970	0,493	0,351	0,649
rs2056277	2	1,64	0,391	0,455	0,545	3	1,660	0,399	0,444	0,556
rs221956	2	1,76	0,432	0,417	0,583	3	1,780	0,439	0,408	0,592
rs2269355	2	2,00	0,500	0,378	0,622	3	2,020	0,505	0,373	0,628
rs2399332	2	1,86	0,462	0,396	0,604	4	2,010	0,502	0,344	0,656
rs279844	2	1,95	0,488	0,408	0,592	2	1,950	0,488	0,408	0,592
rs2830795	2	1,78	0,440	0,425	0,575	4	1,980	0,495	0,348	0,652
rs2920816	2	1,94	0,484	0,355	0,645	3	2,100	0,524	0,296	0,704
rs321198	2	1,83	0,454	0,411	0,589	3	1,840	0,457	0,409	0,591
rs354439	2	1,98	0,494	0,390	0,610	3	2,000	0,499	0,386	0,614
rs430046	2	1,97	0,492	0,359	0,641	2	1,970	0,492	0,359	0,641
rs445251	2	1,98	0,495	0,401	0,599	2	1,980	0,495	0,401	0,599
rs4530059	2	1,93	0,481	0,377	0,623	2	1,930	0,481	0,377	0,623
rs4606077	2	1,49	0,330	0,503	0,497	4	1,600	0,375	0,419	0,581
rs6811238	2	1,98	0,495	0,423	0,577	3	2,000	0,500	0,411	0,589
rs6955448	2	1,43	0,299	0,574	0,426	2	1,430	0,299	0,574	0,426
rs7041158	2	1,73	0,422	0,451	0,550	3	1,750	0,428	0,430	0,570
rs722098	2	1,50	0,334	0,499	0,501	3	1,520	0,343	0,492	0,508
rs727811	2	2,00	0,499	0,397	0,603	2	2,000	0,499	0,397	0,603
rs740910	2	1,73	0,422	0,441	0,559	3	2,100	0,523	0,321	0,679
rs8078417	2	1,82	0,450	0,395	0,605	4	1,860	0,461	0,383	0,617
rs826472	2	1,91	0,476	0,380	0,620	3	1,930	0,483	0,372	0,628
rs876724	2	1,83	0,455	0,396	0,604	3	2,730	0,634	0,206	0,794
rs907100	2	1,99	0,498	0,367	0,633	3	2,470	0,596	0,233	0,767
rs987640	2	1,94	0,485	0,418	0,583	3	2,160	0,536	0,321	0,679
rs9905977	2	1,87	0,464	0,432	0,568	5	2,630	0,620	0,244	0,756
rs9951171	2	1,97	0,493	0,429	0,571	3	2,070	0,516	0,388	0,612

Conclusion

In this study, additional SNPs were detected in the flanking regions of 32 of the 94 iiSNPs genotyped in a population of 93 subjects from North-East Italy.

Variations in A_e, H_e, and PD were observed in 81.25% of the 32 iiSNPs, which became more significant as the MH and mMH frequencies increased.

However, the finding of a haplotype with a frequency <0.01, despite not significantly affecting the forensic parameters of the iiSNP, is not irrelevant, as it contributes to better characterizing the populations and, according to forensic genetic investigations needings, could represent crucial evidence in solving the forensic casework.

Conflict of interest statement

None.

References

[1] K.K. Kidd, A.J. Pakstis, W.C. Speed, et al., Microhaplotype loci are a powerful new type of forensic marker, Forensic Sci. Int. Genet. Suppl. Ser. 4 (2013) e123-e124.

[2] J.L. King, J.D. Churchill, N.M. Novroski, et al., Increasing the discrimination power of ancestry- and identity-informative SNP loci within the ForenSeq™ DNA Signature Prep Kit, Forensic Sci Int Genet. 36 (2018) 60-76.

[3] K.M. Kiesler, L.A. Borsuk, C.R. Steffen, et al., US Population Data for 94 Identity-Informative SNP Loci, Genes 14 (2023) 1071.

[4] L. Davenport, L. Devesse, D. Syndercombe Court, D. Ballard, Forensic identity SNPs: Characterisation of flanking region variation using massively parallel sequencing, Forensic Sci Int Genet. 64 (2023) 102847.

[5] Verogen (2022) ForenSeq™ DNA Signature Prep Reference Guide. Document # VD2018005 Rev. D.

[6] Verogen (2021) MiSeq FGx™ Sequencing System Reference Guide. Document # VD2018006 Rev. F

[7] A.E. Woerner, J.L. King, B. Budowle, Fast STR allele identification with STRait Razor 3.0, Forensic Sci Int Genet. 30 (2017) 18-23.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 327-332
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Forensic Characteristics Y-STRs In the Serbian Population Using The Novel 26 RT Y-STRs Multiplex Assay

Dragana Zgonjanin[1,2*], Reem Lootah[3], Reem Almheiri[3], Khadija Alnahdi[3]
Rashed Alghafri[3]

[1]*Institute of Forensic Medicine, Clinical Center of Vojvodina, Novi Sad, Serbia*
[2]*Faculty of Medicine, University of Novi Sad, Novi Sad, Serbia*
[3]*General Department of Forensic Sciences and Criminology, Dubai Police G.H.Q., Dubai, United Arab Emirates*

** Corresponding author at: Institute of Forensic Medicine, Clinical Center of Vojvodina, Novi Sad, Serbia, Tel.: +381 63 590 686, E-mail address: dragana.zgonjanin.bosic@gmail.com (D. Zgonjanin) dragana.zgonjanin-bosic@mf.uns.ac.rs (D. Zgonjanin)*

Abstract

Recent research with Rapidly Mutating Y-STRs (RM Y-STRs) have shown that these loci provide substantially higher haplotype diversity and haplotype discrimination capacity in worldwide populations when compared with the Y-STRs commonly used in genetic forensics. The discovery of these markers happened in two stages, first stage was in 2010 when 13 markers were discovered, whereas the second stage was the reveal of 12 additional RM Y-STRs in 2021. In our earlier studies, we have estimated a mutation rate for 13 RM-YSTR in 85 pairs of male relatives in the population of Serbia (74 father-son pairs, and 11 twin pairs) to evaluate the capacity of distinguishing between male subjects within a single lineage. In this study, a total of 210 male individuals' samples from population of Serbia were analyzed using 26 RM-Yplex. The assay was developed and validated to amplify 26 markers on Y-Chromosome consisting of 17 single-allelic markers and 10 multi-allelic markers, brining the total number of amplified alleles to 40. The same set of samples was analysed using previously designed 13 RM Y-STR multiplex assay. Therefore in this study we will demonstrate the impact of the additionally released 12 RM Y-STRs compared to the previously revealed 13 TM-Y-STRs and also we will be able to assess the efficiency of the newly designed 26 RM-Yplex assay in differentiating close male relatives based on the analysis of the 26 RM Y-STRs simultaneously. Our findings are encouraging and concur with previous studies showing that by RM Y-STR typing the discrimination power of close male relatives could be considerably increased in comparison to every Y-STR markers commonly used in forensic genetics.

Keywords
Rapidly Mutating, Y-STRs, Haplotype diversity, Forensic genetics, Population of Serbia.

Introduction

Rapidly mutating Y-chromosomal short tandem repeats (RM Y-STRs) were identified to improve differentiation of unrelated males and also to enable separating closely and distantly related males in human identity testing in forensic applications [1-3]. In 2012, Ballantyne described 13 RM Y-STRs showing remarkably increased discrimination power between male lineages in 51 worldwide populations [2]. In 2015, Alghafri introduced a novel system incorporating all those markers in a single multiplex assay kit [4]. Based on the results in our earlier studies [5,6] using that kit, the RM Y-STR loci showed remarkable haplotype resolution power in the population of Serbia, high genetics diversity and, therefore, demonstrating their usefulness in forensic identification cases. In 2020, 12 additional RM Y-STR markers with increased mutation rates were identified by Ralf et al. [7].

Recently, a novel multiplex has been developed and optimized for the simultaneous amplification of all 26 RM Y-STR; 26 RM Yplex [8]. The aim of this novel system is to create a tool that possesses the highest discrimination power between closely related male individuals through the use of 26 RM Y-STRs. This kit has been validated according to the references on forensic analysis as recommended by the International of Forensic Genetics (ISFG) and Scientific Working Group on DNA Analysis Methods (SWGDAM).

In this study, we used this RM-Yplex assay [8] to type 210 unrelated individuals from Serbia. Serbia, with the total surface area of 88.361 km2 and population of approximately 7.5 million, is located on the Balkan Peninsula i.e. in southeast Europe (around 79% of the territory) and in the Pannonian Plain i.e. Central Europe (about 20% of the territory). Genetic diversity indexes and haplotype frequencies were evaluated and the data was used to establish forensic parameters for the population of Serbia.

Material studied, methods, techniques

DNA Samples and extraction

DNA was obtained from bloodstains from 210 unrelated healthy male individuals from Serbia, following informed consent. A 0.5mm disc of blood samples deposited on FTA™ cards was punched out of each male individual sample. The

disc then was washed twice using FTA™ buffer. After the second wash, samples were kept to dry at room temperature for 15 minutes before proceeding by adding PCR reagents.

PCR amplification and typing

PCR amplification was carried out using Veriti 96-Well Thermal Cycler (Thermo Fisher Scientific). The total reaction volume was set to 10 µl, which included 4 µl of Microreader 2.5 × Master Mix I, 2 µl of Microreader RM-Y 5 × Primer Mix, 2 ng of sample DNA, and nuclease-free water to compensate the total 10 µl reaction volume. The parameters for thermal cycling were set as follows: preincubation at 95°C for 5 min, 28 cycles of denaturation at 94°C for 20 s, and annealing at 60°C for 90 s, followed by a final extension at 60°C for 1 h and then holding at 4°C until the amplicons were removed from the thermal cycler instrument. Prior to electrophoresis, 1 µl of the amplified products were added to 0.5 µl Microreader size standard QD650 (Bejing Microreader Genetics, China) and 8.5 µl of deionized Hi-DiTM formamide and denatured for 3 min at 95°C. Amplified DNAs were electrophoresed on an ABI 3500 Genetic Analyzer (Thermo Fisher Scientific) using 36 cm capillaries and POP-4 polymer (Thermo Fisher Scientific) and the fragment analysis was performed with a GeneMapper® IDX v.1.5 (Thermo Fisher Scientific).

Statistical analysis

Forensic parameters were calculated for the 210 typed samples and were compared with previously obtained results for a subset of the same samples (203 samples) [9] for the three former Y-STR marker sets minimal haplotype (MHT, 9 loci), SWGDAM (11 loci), and Yfiler marker panels (17 loci), Yfiler® Plus kit, respectively (Table 1).

Haplotype frequencies and allele frequencies were calculated using the counting method. The genetic diversities (GD) of the markers were calculated according to Nei [11]. The haplotype diversities (HD) were calculated using the formula $HD = n(1 - \sum pi2)/(n - 1)$, where n is the sample size and pi the frequency of the i-th haplotype. The match probability (MP) was calculated as the sum of the squared haplotype frequencies, and Discrimination Capacity (DC) was calculated as the ratio between the number of different haplotypes and the total number of haplotypes. All the above diversity indexes were computed using the Arlequin software ver. 3.5.1.3 [12].

Results

A total of 213 alleles were detected at 26 Y-STR loci and allele frequencies ranged from 0.0048 to 0.4320. The total number of alleles at each locus ranged from 6 for DYS526a to 51 for DYF1000. Nine RM Y-STR loci (DYF387S1, DYF399S1, DYF403S1a, DYF404S1, DYF1001, DYS724, DYR88, DYF1000 and DYF1002) most often show multiple alleles (between one and five alleles), and are therefore categorized as multi copy loci. The inclusion of multi-allelic markers in Y-STR analysis always adding discrimination power which helps narrowing suspected individuals and help excluding innocent individuals from being contributors to biological sample in question.

As previously determined in other populations, the RM Y-STR loci showed high genetics diversity (GD) values (>0.72) in the Serbian population. The highest GD was observed for the locus DYF1000 (0.993), followed by loci DYF1001 (0.992), DYF399S1 (0.990), DYF1002 (0.963) and DYS403S1a (0.956). The lowest GD value was detected for DYS576 which is one of the very first discovered RM Y-STR markers.

Microvariant alleles were rarely observed, except in DYF399S1, DYF404S1 and DYF403S1a. One null allele, in DYF403S1b, was found. Null alleles at this locus have been previously detected in the Dutch population [13].

A total of 210 different haplotypes were identified from 210 unrelated male individuals. The overall haplotype diversity was 1 with a discrimination capacity of 100 (Table 1).

Table 1. Forensic parameters for population of Serbia.

	Serbia (203 samples)					Serbia (210 samples)
	MHT	SWGDAM	Yfiler®	Yfiler® Plus	RM-Yplex	RM-Yplex
Number of haplotypes	121	152	186	200	210	210
Sample size	203	203	203	203	209	210
Number of unique haplotypes	102	125	173	197	208	210
HD	0.98922	0.99351	0.99892	0.99985	0.99998	1
DC (%)	77.86	82.24	93.01	98.50	99.01	100

MHT – minimal haplotype (8 loci); SWGDAM – haplotype of the Scientific Working Group for DNA Analysis-Methods (11 loci); Yfiler–AmpFISTRYfiler (16 loci); YfilerPlus–AmpFISTRYfiler Plus (25 loci); HD - Haplotype diversity; DC – Discrimination Capacity

Discussion

The average gene diversity was high in Serbia (0.86). The gene diversities of all loci were following the same pattern found in worldwide populations when compared to previously published results [3]. In general, the highest gene diversity among all loci was estimated for DYF1000 in the Serbian population.

In this assay, the DYF403S1 is divided into three segments, which each is amplified using different pairs of primers. It was notable that the diversity of total segments included in the current research was recorded as 0.984 whereas in previous research the value of the three alleles detected with one pair of primers is less. This is because of the ability of the new approach to detect an overlapping alleles which were shaded by single peak in the electropherogram.

All 210 were individualized using only 5 Loci including: DYF1000, DYF1001, DYF399S1, DYF1002 and DYS403S1a. Novel abnormal allelic patterns were revealed in Serbian Populations as shown in results.

Conclusion

In this study, we represent the efficiency of Rapidly Mutating Y-STRs markers in discriminating male in Serbian population which eventually leads to great power for discriminating close male relatives.

The study has shown that locus DYF1000, is the mostly diverse Y-STR marker in the population, although it is a multi-allelic marker, it was still showing higher GD value than the highest previously noted Y-STR marker DYF399S1.

Conflict of interest statement

None.

References

[1] K.N. Ballantyne, M. Goedbloed, R. Fang, et al., Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications, Am. J. Hum. Genet. 87 (2010) 341–353.

[2] K.N. Ballantyne, V. Keerl, A. Wollstein, et al., A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages, Forensic Sci. Int. Genet. 6 (2012) 208–218.

[3] K.N. Ballantyne, A. Ralf, R. Aboukhalid, et al., Toward male individualization with rapidly mutating Y-chromosomal short tandem repeats, Hum. Mutat. 35 (8) (2014) 1021–1032.

[4] R. Alghafri, W. Goodwin, A. Ralf, et al., A novel multiplex assay for simultaneously analysing 13 rapidly mutating Y-STRs, Forensic Sci. Int. Genet.17 (2015) 91–98.

[5] Zgonjanin D, R. Alghafri, M. Antov, et al., Rapidly mutating Y-STRs population data in the population of Serbia and haplotype probability assessment for forensic purposes, Forensic Science International: Genetics Supplement Series. 6, (2017) e383-e384.

[6] Zgonjanin D, R. Alghafri, R. Almheiri, et al., Mutation rate at 13 rapidly mutating Y-STR loci in the population of Serbia, Forensic Science International: Genetics Supplement Series. 6, (2017), e377-e379.

[7] A. Ralf, D. Lubach, N. Kousouri, et al., Identification and characterization of novel rapidly mutating Y-chromosomal short tandem repeat markers. Hum. Mutat.41 (9) (2020) 1680–1696.

[8] A. Albastaki, M. Naji, R. Lootah, et al., Design and development of novel single multiplex system incorporating 26 rapidly mutating Y-STRs; 26 RM Yplex. Electrophoresis 43 (2022) 1911–1919.

[9] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing for forensic material, Biotechniques 10 (1991) 506–513.

[10] D. Zgonjanin, R. Alghafri, M. Antov, et al., Genetic characterization of 27 Y-STR loci with the Yfiler Plus kit in the population of Serbia, Forensic Science International: Genetics 31 (2017) e48-e49.

[11] M. Nei, F. Tajima, DNA polymorphism detectable by restriction endonucleases, Genetics 97 (1981) 145–163.

[12] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Res. 10 (2010) 564–567.

[13] A. A. Westen, T. Kraaijenbrink, L. Clarisse, et al., Analysis of 36 Y-STR marker units including a concordance study among 2085 Dutch males, Forensic Sci. Int. Genet. 14 (2015) 174–181.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 333-339
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Optimizing Template Input: Complex Family pedigrees M-FISys with LIMS

Diana C. Vinueza-Espinosa[1*], María Isabel Medina-Lozano[1], María Sáiz Guinaldo[1], Xiomara Gálvez Escolano[1], Christian Haarkötter Cardoso[1], José Antonio Lorente[1], Juan Carlos Álvarez Merino[1]

[1]*Laboratory of Genetic Identification & Human Rights (LABIGÉN), Department of Legal Medicine, Toxicology & Physical Anthropology, Faculty of Medicine, University of Granada (UGR)*

** Corresponding author at:*
E-mail address: dvinuezaespinosa@ugr.es, (D. Vinueza-Espinosa)

Abstract

Various databases and tools are currently used to facilitate the identification of unidentified human remains (UHRs) through direct or kinship DNA matching. M-FISys (Mass-Fatality Identification System), originally created to identify victims from the 9/11 World Trade Center attack, supports both direct and indirect DNA profile matching. In our laboratory, the focus is on identifying victims from the Spanish Civil War (1936–1939) and postwar Andalusian victims, primarily through kinship matching using M-FISys. For skeletal remains, kinship matching involves inputting a family pedigree, allowing the system to search for samples that match all family members and compute a joint probability likelihood ratio (JPLR) to indicate the likelihood of a match. M-FISys constructs these pedigrees through two different templates: the Basic Family template and the Complex Family template. The Basic Family template automatically generates pedigrees using pre-established relationships such as parent, sibling, etc., making it straightforward to use. However, the Complex Family template requires manual input of specific details, including family codes and donor data. Although this method allows for more customised entries, it is labour-intensive and prone to human error.

To streamline this process and reduce errors, we integrated a Laboratory Information Management System (LIMS) and employed QuickPed software to semi-automate the creation of pedigrees, generating customised lab codes for each family group. The optimisation allowed for reducing the time-consuming and prone to critical errors that can affect profile matching, moreover the export of files that are ready for direct import into M-FISys, containing all relevant information, including genetic profiles. Furthermore, this approach guarantees complete traceability of

the samples, from the initial DNA extraction to the final genetic profile, a key requirement for laboratories working towards ISO/IEC 17025 accreditation.

Keywords
M-FISys, Kinship, Genetic Identification, Family Complex, QuickPed, Skeletal Remains.

Introduction

Currently, various databases and software tools assist analysts in using reference DNA to identify unidentified human remains (UHRs) through Direct or Kinship matching. M-FISys (Mass-Fatality Identification System) employs direct and indirect profile matching, originally designed for identifying remains from the 9/11 attack on NYC's World Trade Center [1]. Our lab specializes in identifying victims of the Spanish Civil War (1936–1939) and postwar Andalusian victims. The identification of skeletal remains from the Spanish Civil War and postwar is conducted through kinship matching using M-FISys.

Given a pedigree of family members for one or more missing persons, the database can be searched for samples that match all family members, to calculate the Joint Probability Likelihood Ratio (JPLR), which indicates the probability that the match is related to all family members as presented. M-FISys constructs pedigrees using relationships specified from input files. Two templates are used in M-FISys: Basic Family and Complex Family (Figure 1). The Basic Family template, which requires no effort, automatically draws pedigrees using relationships defined in the template, such as father, mother, sibling, etc. The Complex Family template involves defining parents for each member, with columns including Family Code, Father/Mother Fname, Lname, and Donor Code. Filling out the Complex Family template manually is time-consuming and error prone.

We have implemented a semi-automated way to complete the complex template with personalized lab codes for pedigrees using a Laboratory Information Management System (LIMS) along with the free online software QuickPed [2].

Material studied, methods, techniques

A total of 804 pedigrees were created using the QuickPed tool [2] from 16 kinship cases. These kinship cases included relationships such as parents, siblings, uncles, cousins, second cousins, grandparents, great-grandparents, and great-great-grand-

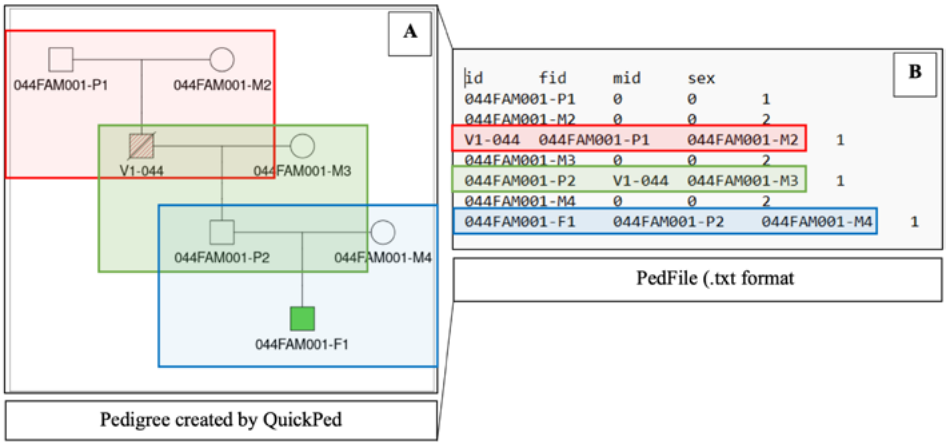


Figure 1. Example of a pedigree created using the Quickped tool, where the family member is searching for one victim (the parental grandfather). A. graphic representation of the pedigree. B. Text representation of the pedigree.

parents, where reference family was searching for one or two victims. The pedfile (in.txt format) was downloaded from the QuickPed tool; this file contains information on all family members, allowing them to be linked together and to establish a specific relationship with personalised lab codes (Figure 1). In the LIMS, the data for family reference (e.g., 044FAM001-F1, green square) and the victim (e.g., V1-044, lined square) were registered as items in each case. The family references were subsequently uploaded as samples, having been technically analysed to obtaind a genetic profile. The family group (pedigree) was also entered into the LIMS, within a pre-designed environment (programming code) that includes fields (such as "*Father and mother donor code*"). to be completed in the family complex template in order to construct the pedigree in M-FIsys databases. The information entered into LIMS was exported in the family complex template format (.xls format) and imported into M-FIsys. Errors were analysed by counting them in two files.

The first set of errors, found in the export *file.xls* generated by LIMS, were classified as LIMS errors (LEs), and the second set, listed in the validation report generated during the import into the database, were classified as M-FIsys errors (MEs). The M-FIsys database includes a validator to must be run before importing the LIMS *xls file* (save as *txttab*) into the M-FIsys database.

The LEs were categorised into two distinct types (Table 1). In LIMS, the *pedfile* information was entered by coping and pasting the family member's codes (green, red and blue highlighted boxes in figure 1), which allows them to be linked

to one another in order to establish the specified relationships. If the codes were copied and pasted incorrectly or if family members were missing, these errors were classified as manual errors (Type of LE: 1). Type of LE:2 included issues arising from the exportation of the LIMS software, due to programming code problems or warnings that occurred during its development.

Table 1: Types of LIMS errors evaluated in the file to be imported to M-Fisys database

Type of LE	Observations
1	manual errors
2	internal of the software

The MEs were categorized into four different types (Table 2). These errors were identified through the M-FIsys Validator report and an additional report generated during the file import process into the database.

Table 2: Types of M-FIsys errors evaluated in the file before importing into the M-FIsys database

Type of ME	Observations
1	file format error
2	warnings
3	data errors
4	code repetition

Results and Discussion

Results indicated that there were generally more manual errors (type 1) than LIMS-internal (type 2) (Figure 2A). This fact is likely due to the initial challenges faced by analysts in adapting to the copy-and-paste method during the early stages of implementing this semi-automated process. Furthermore, the task was initially carried out by different analysts, which contributed to the variation in errors. However, these manual errors were significantly reduced when the task was assigned to a single analyst.

Regarding the MEs, the results showed that there were more warning errors (type 2), followed by data errors (type 3), and with the lowest number of errors, type 1 (file format error) and 4 (code repetition) (Figure 2B).

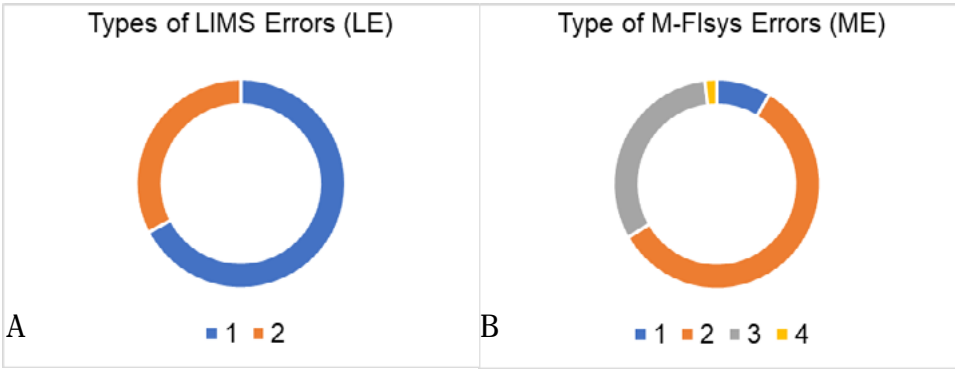
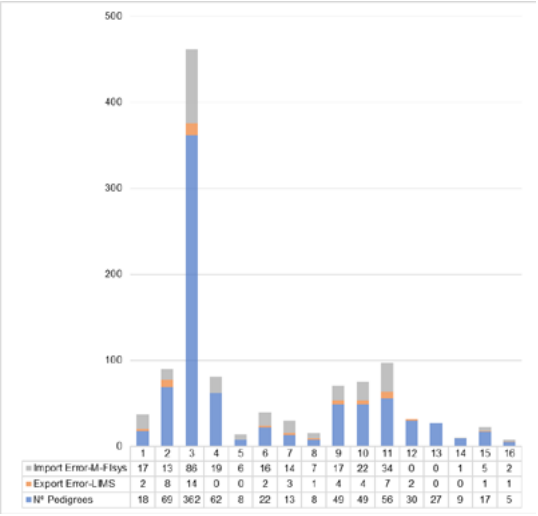


Figure 2. (A) Types of LIMS errors-LE. **(B)** Types of M-FIsys errors evaluated in the file to be imported to M-FIsys database.

In the 804 pedigrees evaluated across 16 kinship cases, most errors occurred during the file import process into the M-FIsys database. However, the majority of these errors were classified as warnings (Figure 3). Warning errors typically arise when a particular allele is not present in the frequency table, triggering a warning because the minimum frequency is applied. These warnings are not critical and do not affect the kinship match in the database.

Another frequently encountered issue was the data error. This occurred when the "Assay kit" field in the family complex template contained molecular marker information that was not applicable to the selected kit. For example, in a sample analysed using the "*Global Filer*" assay kit, the molecular marker "Penta E" was mistakenly assigned a value of "0"in the template. When this file was imported into M-FIsys, a data error was flagged. This occurred because, in kits like GlobalFiler, cells corresponding to molecular markers not included in the kit should remain empty, without a "0" or any other value. This error was easily corrected by updating the LIMS programming code to ensure that fields for molecular markers not relevant to the selected assay kit remain blank.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Import Error-M-FIsys	17	13	86	19	6	16	14	7	17	22	34	0	0	1	5	2
Export Error-LIMS	2	8	14	0	0	2	3	1	4	4	7	2	0	0	1	1
N° Pedigrees	18	69	362	62	8	22	13	8	49	49	56	30	27	9	17	5

Figure 3. Errors evaluations of the template input exported from LIMS and imported to M-FIsys database in a total de 804 pedigrees from 16 kinship cases.

Conclusion

In conclusion, the minor and easily detectable errors flagged by the M-Fisys data validator do not impact the accurate entry of pedigrees into the database, making the process highly efficient and significantly reducing the time required for genetic comparison analysis. This automated approach provides a clear advantage over the manual method of completing the.xls template, which is both more time-consuming and prone to critical errors that can affect profile matching. Furthermore, this optimatisation allows for seamless export of ready-to-import files from LIMS to M-FISys, ensuring full traceability of samples from DNA extraction to the generation of genetic profiles -an essential factor for laboratories pursuing ISO/IEC 17025 accreditation.

Acknowledgments

Special Thanks to LabWare, particularly Sandra Gonzalez Granado, for her invaluable support during the challenging development process of linking M-FIsys template with LIMS, and for her assistance in creating our tool for constructing family groups (pedigrees). Likewise, Howard Cash and Mike Hennessey from Gene Codes

Corporation are acknowledged for their support in the training for M-FISys database.

Conflict of interest statement

None

References

[1] H.D. Cash, J.W. Hoyle, A.J. Sutton, Development under extreme conditions: forensic bioinformatics in the wake of the World Trade Center disaster, Pac.Symp. Biocomput. (2003) 638–653.

[2] M.D. Vigeland, QuickPed: an online tool for drawing pedigrees and analysing relatedness, BMC Bioinformatics. 23 (2022).

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 341-347
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Use of 3915 kinship SNPs-Microarrays in missing person identification in the absence of two generations

Sebastian Biagini[1], Federico Picado[1], Santiago Ginart[1], Malena Canteros[1], Valeria Genoud[1], Florencia Gagliardi[1], Nicolas Furman[1], Constanza Mayordomo[1#], Mariana Herrera Piñero[1#*].

[1] *Banco Nacional de Datos Genéticos (BNDG), Av. Córdoba 831, CABA, Argentina.*

E-mail: info@bndg.gob.ar, Tel.: +54 11 4891 8951

[#] *Both authors contributed equally*

[*] *Corresponding author: Mariana Herrera Piñero, mherrera@bndg.gob.ar*

Abstract

In missing person identification, the statistical power derived from genetic STR profiles of a family group is crucial for distinguishing the person of interest from unrelated individuals. Specifically, in cases involving only second and/or third-degree relatives, the statistical power is significantly reduced, and STR markers often fail to provide sufficient discriminatory information. Our primary goal is to validate the enhanced discrimination power in complex cases using SNP profiles obtained by microarray technology. We accomplish this by first compiling a comprehensive frequency table of the Argentine population based on 3,915 SNPs. Subsequently, we analyze the statistical power of identification for missing grandchildren and great-grandchildren utilizing these markers. This approach aims to address the challenges posed by incomplete family groups in solving complex kinship cases. To assess our aims, we conducted a comprehensive analysis of 288 whole-blood samples using the salting-out protocol, adhering to the purity and integrity guidelines recommended by the manufacturer. We employed the Illumina Infinium Global Screening Array-24 Kit to perform SNP marker typing via microarray technology. From the resulting dataset, we selected 3,915 unique SNPs previously associated with known kinship relationships. We then utilized this data set to simulate conditional scenarios based on three previously resolved STR cases. For this simulation, we applied Merlin software to calculate likelihood ratios for various hypotheses, focusing on one and two pedigree members of second and third-degree relatives. In the present study, we present the first population frequency table of Argentina for kinship SNPs markers obtained by microarrays technology. Furthermore, we demonstrated that the selected panel of 3,915 kinship SNPs effectively minimizes false positive and

false negative rates and validated this method for complex cases such as cases of crimes against humanity, when only one second or third-degree relative of the sought individual is available.

Keywords
Microarray, Kinship SNPs, DVI-MPI, Linkage, Allele Frequency Table, Simulation.

Introduction

In missing person identification, the statistical power of genetic STR profiles from family members is essential for distinguishing the person of interest from unrelated individuals. However, when only second- and third-degree relatives are available, the statistical power decreases significantly, and STR markers alone become insufficient. Incorporating single nucleotide polymorphism (SNP) analysis proves effective in determining distant biological relationships. Advances in technology, such as microarray techniques, have made it possible to obtain these results more efficiently [1, 2].

The Banco Nacional de Datos Genéticos was a pioneer in conducting genetic studies to address crimes against humanity, such as forced disappearances and the restoration of identity for children appropriated during the last dictatorial regime (1976-1983) that occurred in Argentina. It maintains a searchable database that enables large-scale comparisons between individuals of interest (such as grandchildren) and extended family groups, even when genetic information about their missing parents is unavailable [3, 4].

The objective of this work is to create a frequency table of the Argentine population based on 3,915 SNPs and analyze the statistical power of identification for missing grandchildren and great-grandchildren using these SNP markers to solve complex cases involving incomplete family groups.

Materials and methods

We analyzed the genotype of 288 blood samples from Argentinian donors. All the samples provided were analyzed in the BNDG within Law 26.548 and accompanied by a signed consent form accepting the use of their genetic profiles for population studies.

DNA from blood samples collected in EDTA tubes was extracted using Salting Out protocol and quantified by DS-11 DNA Quantification Spectrophotometer (De-

novix). DNA sample quality was assessed by reading the absorbances 260/280 and 260/230. Qubit 4 fluorometer (Thermo Fisher) was used to determine the integrity and quantity of the sample obtained [5]. For all samples, a 200 ng of template DNA according to the manufacturer's microarray protocol were obtained. SNP typing was performed using microarray technology with the Illumina Infinium Global Screening Array-24 Kit (Illumina, San Diego, USA) [6].

All genotype assignments met the required read quality parameters associated with Hardy-Weinberg equilibrium, obtained by using GenomeStudio software, following the recommendation of Guo et al. 2014 [7] and the Illumina's guidelines [8]. Based on the FORCE panel designed by Tillmar et al. [9], we selected 3,915 SNPs that are relevant for kinship analysis. To build the first kinship SNPs frequency table for the Argentine population, 181 samples from unrelated individuals were analyzed (Data not shown). For the allelic assignment we use the Illumina annotation criteria [10]. Additionally, the 3,915 SNPs selected for the frequency table were recalculated by using Bonferroni correction [7], and all were found to be in Hardy-Weinberg equilibrium.

Three previously solved cases with STRs were used for conditional simulations based on the SNPs profiles. The statistical analyses were performed using R v.4.2.0. Pedigree management and simulations were done with pedsuite packages collection [11] and mispitools package (https://cran.r-project.org/web/packages/mispitools/). For the LR distribution we performed 1000 simulations and calculations of the likelihood ratios (LRs) for the related and unrelated Missing Person (MP) for the tested hypotheses [12]. To do this we selected 1 and 2 pedigree members of second and/or third-degree relatives. To calculate the linkage LR, the kinshipLR function from the forrel package and simLRgen function from the mispitools package were adapted to account for the use of the MERLIN software [13].

Results and discussion

To analyze the general statistical power of the selected 3,915 kinship SNP markers, we built a generic pedigree and performed simulations based on the genotypes of a duo (Fig.1 left). Using our Argentinean SNP frequency table (Data not shown), we evaluated the linkage Likelihood Ratio (LR) in a duo comparison between MP and second- and third-degree relatives (Fig.1 right). The distribution plots show a clear separation between related and unrelated individuals across all tested relationships, allowing for the confident identification of the MP by minimizing the

false positive and false negative rates. Therefore, we were able to validate this in our family groups' caseworks. These kinship SNPs, previously described in the FORCE Panel [9], allow us to achieve high statistical power in identifying a missing person when only a second- or third-degree relative is available, as reported similarly by Tillmar et al. [9].

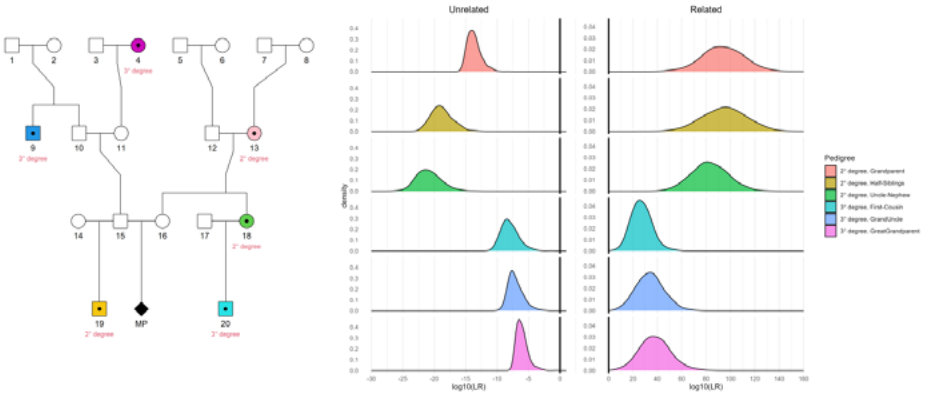


Fig. 1 Generic pedigree plot and linkage LR distribution plot for member's combination with MP.

The pedigree plot (left) illustrates the generic family structure, highlighting the relatives' pedigree to simulate. The linkage LR distribution plot for unrelated and related individuals (right) where the colors correspond to different pedigree members. The simulations were plotted on separate graphs because the density scale on the Y-axis impaired the visualization of related individuals.

To test our simulated outcomes with SNP profile data, we genotyped the pedigree member's for two previously solved identifications with STR. FAM1 (Fig. 2A) have one incomplete generation and are searching for a child who is the MP. Therefore, combinations of second- and third-degree relatives were used to perform conditional LR simulations and compared them with the true LR (linkage and no linkage) value obtained from the inclusion. We observe a clear distinction between related and unrelated individuals for all combinations. The true LR value (black diamond) for the identified MP lies within the estimated LR values for various combinations of pedigree members.

To advance our analysis, we evaluated the possibility of two incomplete generations by using combinations of third-degree related individuals for conditional LR simulations in the case of FAM2 (Fig. 2B), which includes three biological

children from the identified MP (the great grandchildren). Under these conditions, we observe a good separation between related and unrelated individuals. Despite having only third-degree relatives and missing two generations, accurate identification can still be achieved for all three cases (indicated by the black diamonds).

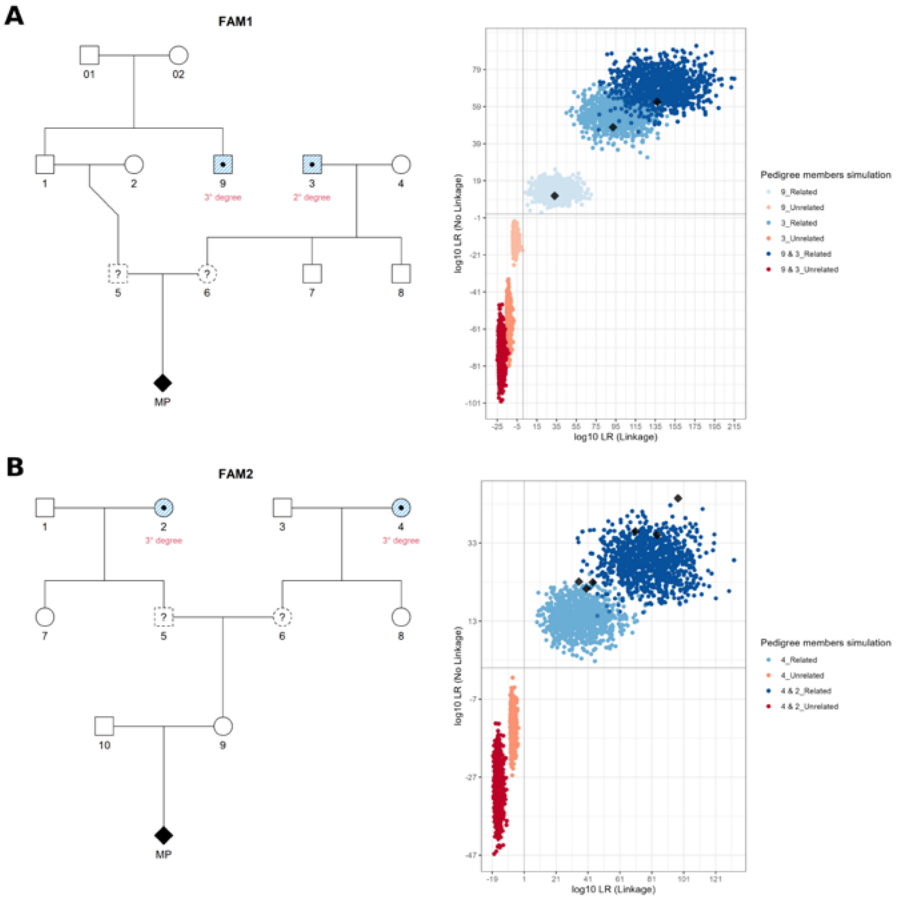


Fig. 2. Linkage and No Linkage LR Distributions for Pedigree Members.

Simulation results in absence of one generation FAM1 (A); absence of two generations FAM2 (B). The pedigree plot (left) illustrates the family structure, highlighting second- and/or third-degree relatives. The scatter plot (right) shows the log10 LR (Linkage) versus log10 LR (No Linkage) distributions for 1000 simulations, categorized by different pedigree member combinations. The clusters indicate the

separation of related (blue tones) and unrelated (red tones) individuals based on the LR calculations. The black diamond represents the true LR value from different pedigree member combinations.

Conclusion

The results obtained from SNP profiles of reference family members in previously resolved STR cases highlight the panel's potential for solving complex missing person identification cases involving second and third-degree relatives, with clear differentiation between related and unrelated individuals. Kinship SNPs markers offer a powerful tool for identifying a MP in incomplete family groups in the absence of information from two generations, thereby enhancing the capabilities of traditional STR-based methods.

Conflict of interest statement

None

References

[1] Divne, AM. and Allen, M. A DNA microarray system for forensic SNP analysis. Forensic science international, 154(2-3) (2005), 111-121.

[2] Voskoboinik, L., Ayers, SB., LeFebvre, AK. *et* al. SNP-microarrays can accurately identify the presence of an individual in complex forensic DNA mixtures. Forensic Science International: Genetics, 16 (2015), 208-215.

[3] Herrera Piñero, M., Stover E., Tupa M. *et* al., 'The Living Disappeared: Forensic DNA Typing and the Search for Argentina's Stolen Children', in Henry Erlich, Eric Stover, and Thomas J. White (eds), Silent Witness: Forensic DNA Evidence in Criminal Investigations and Humanitarian Disasters, online edition, Oxford Academic (2020), doi.org/10.1093/oso/9780190909444.003.0008.

[4] Kling, D., Egeland, T., Piñero, MH., *et* al. Evaluating the statistical power of DNA-based identification, exemplified by 'The missing grandchildren of Argentina'. Forensic Science International: Genetics, 31, (2017), 57-66. doi.org/10.1016/j.fsigen.2017.08.006

[5] Sealfon, SC. and Chu, TT. RNA and DNA Microarrays. In: Khademhosseini, A., Suh, KY., Zourob, M. (eds) Biological Microarrays. Methods in Molecular Biology, vol 671. 2011, Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-59745-551-0_1

[6] Infinium Global Screening Array-24 Kit. https://www.illumina.com/products/by-type/microarray-kits/infinium-global-screening.html

[7] Guo, Y., He, J., Zhao, S. *et* al. Illumina human exome genotyping array clustering and quality control. Nat Protoc 9, (2014), 2643–2662. https://doi.org/10.1038/nprot.2014.174

[8] Custom cluster file creation for improved copy number analysis https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/custom-cluster-file-tech-note-m-gl-02142/custom-cluster-file-tech-note-m-gl-02142.pdf

[9] Tillmar, A., Sturk-Andreaggi, K., Daniels-Higginbotham, J. *et* al. The FORCE panel: an all-in-one SNP marker set for confirming investigative genetic genealogy leads and for general forensic applications. Genes, 12(12), (2021),1968. doi.org/10.3390/genes12121968

[10] "TOP/BOT" Strand and "A/B" Allele https://www.illumina.com/documents/products/technotes/technote_topbot.pdf

[11] Vigeland MD. 2021. Pedigree Analysis in R. Academic Press. ISBN 978-0128244302.

[12] Iungman, M., Biagini, S., Canteros, M. *et* al. Empirical validation of a family-member prioritization approach to maximize statistical power in missing person cases. Forensic Science International: Genetics Supplement Series, 8, (2022), 271-273. doi.org/10.1016/j.fsigss.2022.10.059

[13] Abecasis GR., Cherny SS., Cookson WO. *et*. al. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet.30(1), (2002), 97-101. doi.org/ 10.1038/ng786

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 349-355
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Assessing a High Throughput SNP Assay for Closed Scenario Kinship Investigation

Kevin M. Kiesler[1*], Carolyn R. Steffen[1], Lauren E. Mullen[1], Katherine B. Gettings, and Peter M. Vallone[1]

[1] *National Institute of Standards and Technology, USA.*

** Corresponding author at: NIST, 100 Bureau Drive, Gaithersburg MD 20850 USA*
email: kevin.kiesler@nist.gov phone: +1(301)975-4306

Abstract

Researchers at the U.S. National Institute of Standards and Technology (NIST) have conducted experiments using the ForenSeq Kintelligence-HT kit, a highly multiplexed single nucleotide polymorphism (SNP) panel intended for use in disaster victim identification scenarios. The kit works with the MiSeq FGx DNA Sequencing System to assay more than 10000 SNP loci, with 9867 SNPs for the primary task of kinship estimation, and 106 X-chromosome SNPs, 85 Y-chromosome SNPs, 56 ancestry informative SNPs (aiSNPs), 24 phenotype informative SNPs (piSNPs), and 94 individual identification SNPs (iiSNPs) for additional functionality. The system includes a dedicated Universal Analysis Software (UAS), which offers a local server and an offline database for sample comparison and kinship calculations. The Kintelligence-HT kit is designed to analyze up to 36 high-quality antemortem reference samples or 12 postmortem-type samples per sequencing flowcell, providing throughput needed for disaster scenarios. NIST experimental samples included parent-parent-child trios (n = 2) and pairs of alleged brothers (n = 4) as well as degraded DNA with degradation index ranging from 2.7 to 43.5, and low-input DNA of 100 pg and 500 pg. UAS (v2.6) kinship calculations reflected the relationships between first-degree relatives. Analysis of challenging degraded DNA yielded a slightly lower average SNP locus call rate of 95.5 % (± 1.8 %), relative to non-degraded samples (97.9 % ± 0.3 %) in the same sequencing run. Low DNA-input samples had lower average SNP locus call rates at 100 pg input (84.8 % ± 3.1 %) than samples run with 500 pg input (93.6 % ± 6.5 %) or 1000 pg input (99.3 % ± 0.2 %). Challenging degraded or low input samples were successfully searched in the UAS 2.6 database for kinship analysis. Multiplexing 36 samples on a single flowcell run resulted in an average SNP locus call rate of 88.7 % (± 4.5 %) for kinship markers. Lower SNP call rates may create uncertainty in kinship assessment by suppressing the estimate of shared DNA segments.

Keywords
Single Nucleotide Polymorphism, Disaster Victim Identification, Kinship.

Introduction

Within the last decade, the investigative value of public databases containing genetic information from large numbers of individuals has become evident, with the identification of genetic relatives enabling the investigation of unidentified remains or crime scene evidence with no match in a criminal DNA database. Initially developed as an aid to traditional genealogical research on family history and in anthropology, the use of genetic information (a.k.a. genetic genealogy) has seen tremendous popularity, and several companies offer relatively low-cost DNA analysis for those seeking knowledge about their ancestry and extended family history. The result has been the creation of privately owned databases containing millions of genetic profiles. As this approach proved applicable to crime scene investigation, specific databases were made available to law enforcement (LE) use through the adoption of terms and conditions agreements for consent to LE searching, and the GEDmatch PRO (QIAGEN) database was created expressly for the use of forensic and investigative genetic genealogy (FIGG). The DNA sequencing system described here allows a person's profile of single nucleotide polymorphisms (SNPs) from questioned samples to be directly compared with known family reference samples in an offline local database. To that end, Verogen (a QIAGEN company) has recently introduced the ForenSeq Kintelligence-HT Kit (QIAGEN, Hilden, Germany), which incorporates the SNP marker content from the ForenSeq Kintelligence Kit (QIAGEN) [1] with the addition of a PCR enhancer for difficult-to-amplify postmortem samples and an analysis module in the Universal Analysis Software (UAS) (QIAGEN) which runs a local, offline database for sample comparisons and kinship calculations. The intended use of the kit is for confirmation of postmortem sample identity through kinship analysis in mass disaster investigation and remains association, which often requires a higher throughput of samples than investigative genetic genealogy with crime scene samples. The ForenSeq Kintelligence HT Kit allows for the analysis of 12 postmortem samples or 36 antemortem reference samples in a single sequencing run on the MiSeq FGx Sequencing System (QIAGEN).

Material studied, methods, techniques

Materials and Sample Preparation

An assessment was performed with the Kintelligence-HT using three sequencing runs on the MiSeq FGx Sequencing System. Sequencing Run 1 examined sequencing yield, SNP call rate, and reproducibility of SNP profiles and kinship statistics using three replicate library preparations of each of three DNA input values (100 pg, 500 pg, and 1000 pg) of total genomic DNA extract from cell line NA24385 and three no-template controls. Sequencing Run 2 assessed "postmortem type sample" performance with 12 DNA samples in the run: six degraded DNA simulated 'postmortem' samples and four reference-quality samples. Degraded DNA was created by two methods: 1) Two samples, NIST RM 8392 (Ashkenazi son, "AK son") and NIST RM 8393 (Han son), were treated with ultrasonication in a Covaris S2 Focused Ultrasonicator (Covaris, Woburn MA, USA) in a volume of 130 µL using microTUBE AFA Fiber Snap-Cap tubes (Covaris) with 60 s treatments (Frequency Sweeping mode, duty cycle 10 %, intensity 10, 100 cycles per burst temperature ≈ 6 °C) repeated for total sonication times of four or 15 min. 2) Two samples from NIST Reagent Grade Test Material (RGTM) 10235 (samples two and four) were irradiated by short-wave U.V. light for five min in a PCR enclosure approximately 12 cm from the light source. Artificially degraded DNA was assessed for concentration and degree of degradation by qPCR with the Quantifiler Trio Kit (Thermo Fisher Scientific, Waltham MA, USA) on an Applied Biosystems 7500 Real-time PCR System for Human Identification (Thermo Fisher).

Run 3 contained 36 samples of high-quality 'antemortem' reference DNA, including the parents of degraded samples (AK son and Han son) in Run 2, four alleged brothers of the four non-degraded samples in Run 2, and non-degraded 'self' samples corresponding to the two U.V. treated DNA on Run 2, plus 24 unrelated individuals. Positive and negative amplification controls were included in Runs 2 and 3 for quality monitoring. All libraries were constructed per the manufacturer's recommended procedure [2]. Performance was assessed through read counts mapped per sample, SNP call rates, kinship prediction statistics (kinship coefficient, Kc), and estimates of shared DNA (shared centimorgans, cM).

Data Analysis

DNA sequencing output was analyzed using the UAS (v2.6) to map reads to individual samples and call SNPs using default analysis parameters. Kinship comparisons were performed within the UAS local database, comparing samples with

known relationships and/or unrelated samples. Relatedness between samples tested was limited to first-degree relatives (brother-brother, father-son, mother-son) and self-self. More distant relationships were not tested in this study. A comparison of sequencing metrics was performed using paired t-testing to assess the performance of degraded versus non-degraded samples in Run 2.

Results

Artificially degraded DNA was characterized by degradation index (DI) calculated from Quantifiler Trio, with values ranging from 2.7 to 43.5. Sonication for four min produced moderately degraded DNA with DI of 4.8 (AK son) and 3.4 (Han son), whereas sonication for 15 min produced highly degraded DNA with DI of 42.5 (AK son) and 43.5 (Han son). Treatment with U.V. produced moderately to highly degraded DNA with DI of 2.7 (RGTM 10235 Sample 2) and 6.4 (RGTM 10235 Sample 4).

Template gDNA input to the initial amplification (PCR1) was found to be correlated with total number of reads mapped per sample, with 100 pg input DNA yielding an average of 6.1×10^5 ($\pm 1.0 \times 10^5$), which was lower than that of 500 pg input, at 19×10^5 ($\pm 9.2 \times 10^5$) while 1000 pg DNA input yielded the highest amount of sequencing reads per sample at 37×10^5 ($\pm 3.0 \times 10^5$). Call rates were similarly affected by DNA input amount, with an average of 84.7 % (± 3.1 %) for 100 pg input, 93.3 % (± 6.6 %) for 500 pg input, and 99.3 % (± 0.5 %) with 100 pg DNA input. One of the three replicates of the 500 pg DNA input samples yielded significantly lower sequencing reads for that sample by five orders of magnitude relative to the other two replicates, possibly due to a small amount of residual ethanol in the sample from the bead-based cleanup following the initial PCR amplification, causing inhibition of downstream library preparation steps.

SNP profiles from sensitivity testing replicates were compared against each other to assess genotyping reproducibility. The number of SNP calls shared between samples closely mirrored the SNP call rate, with 79.6 % (± 1.8 %) of kinship SNPs overlapping in 100 pg DNA input replicates, 89.3 % (± 6.3 %) in replicates of 500 pg DNA input, and 99.7 % (± 0.2 %) from 1000 pg DNA input samples.

The degree of sample multiplexing within sequencing a run had an inverse effect on the average sequencing coverage mapped per sample. Run 2 consisted of 12 samples, which had an average of 23×10^5 ($\pm 4.5 \times 10^5$) sequencing reads per sample in comparison to Run 3 with 36 samples, where the average sequencing yield per sample was 11×10^5 ($\pm 1.7 \times 10^5$). Negative controls were excluded from

calculations of average reads per sample. Sonicated degraded DNA yielded similar amounts of sequencing coverage 23 x 10^5 (± 3.8 x 10^5) as non-degraded DNA in the same run at 26 x 10^5 (± 2.0 x 10^5) reads. Paired t-testing showed no statistically significant differences in sequencing coverage between sonicated and reference-quality DNA ($p = 0.34$). There were not enough replicates of U.V. degraded DNA to perform statistical analysis.

Table 1. Average number of reads mapped per category of samples and average kinship call rates. The first three rows come from Run 1. Errors are standard deviations.

	Average Number of Sequencing Reads Per Sample (× 10^5)	Average Kinship SNP Call Rate (%)
100 pg DNA	6.1 ± 1.0	84.7 ± 3.1
500 pg DNA	19 ± 9.2	93.3 ± 6.6
1000 pg DNA	37 ± 3.0	99.3 ± 0.5
Run 2	23 ± 4.5	96.6 ± 1.8
Run 3	11 ± 1.7	91.4 ± 2.7

Coverage per locus exhibited similar trends as overall sequencing coverage. Comparing four representative samples from each of Runs 2 and 3, the average coverage mapped per SNP locus was 260.2 (± 262.4) and 114.4 (± 121.7), respectively. The large standard deviation may be the result of some loci having much higher coverage (highest number of reads mapped: Run 2 = 8378, Run 3 = 4395).

SNP call rates were correlated with sample multiplexing; in Run 2 with 12 samples, 96.6 % (± 1.8 %) of kinship SNPs were successfully genotyped, where in Run 3 the average was 91.4 % (± 2.7 %) reflecting the effect of having more samples in the sequencing run. SNP call rates for sonicated degraded DNA were not significantly different from those of reference quality DNA in Run 2 ($p = 0.1695$) using paired t-testing.

Sample to sample comparisons for 1st degree relatives returned consistent kinship coefficient values with an average of 0.26 (± 0.01), very near the theoretical expectation of 0.25. Self-comparisons (degraded postmortem vs. non-degraded antemortem reference) returned kinship coefficients of 0.52 (± 0.00), where Kc above 0.49 is the expected value. Estimates of shared DNA (in cM) averaged 2672.6 (± 315.0) for brother pairs. Father-son and mother-son comparisons were less variable

at 3210.7 (± 33.6) for the AK son and 3094.0 (± 42.4) for the Han son, where reference samples were compared with both moderately and highly degraded versions of the son samples.

Discussion

While it was observed that sequencing yield during sensitivity testing was correlated to DNA input, the number of SNP loci called was greater than the required 6,000 genotypes (corresponding to a 60.8 % call rate of kinship SNPs) to qualify for submission to GEDMatch Pro, a large database of genetic profiles consented for LE use in investigative genetic genealogy. The UAS v2.6 software contains a recommendation of 700000 as a minimum number of sequencing reads per sample. In this study, meeting that recommendation resulted in a greater than 80 % complete profile for sample comparison within the UAS local database. However, as the number of successfully genotyped SNP loci decreases, so does the estimate of shared centimorgans. It may be assumed that profiles with diminished numbers of SNPs may appear to be more distantly related than ground truth relationships.

Conclusion

In this assessment, it was observed that the sequencing yield per sample was proportional to the template gDNA input for PCR amplification and that 100 pg of gDNA was sufficient to generate an adequate SNP profile for comparisons. Lower DNA input amounts were not tested. The degree of sample multiplexing was also inversely correlated to sequencing yield per sample. Artificially degraded DNA exhibiting extreme degradation index values functioned equally as well as reference quality DNA, a notable advantage of using SNP markers with compact amplicon design. In the framework of disaster victim identification, where environmental conditions may produce very challenging samples, utility with degraded DNA can offer significant advantages over archetypal capillary electrophoresis-based methods using short tandem repeat markers.

Acknowledgments

Sincere gratitude to QIAGEN R&D, Melissa Kotkin, and Rebecca Schütt for their invaluable assistance in facilitating this work. This project was supported in part by the NIST Special Programs Office: Forensic Genetics and by an interagency

agreement with the U.S. Federal Bureau of Investigation: DNA as a biometric. All work presented in this paper has been reviewed and approved by the NIST Research Protections Office under protocol # MML-16-0080.

Conflict of interest statement

The authors have no conflicts of interest to declare. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. Certain commercial software, instruments, and materials are identified to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

All work has been reviewed and approved by the National Institute of Standards and Technology Research Protections Office. This study was determined to be "not human subjects research" (often referred to as research not involving human subjects) as defined in U. S. Department of Commerce Regulations, 15 CFR 27, also known as the Common Rule (45 CFR 46, Subpart A), for the Protection of Human Subjects and therefore not subject to oversight by the NIST Institutional Review Board.

References

[1] J. Antunes, P. Walichiewicz, E. Forouzmand, R. Barta, M. Didier, Y. Han, J. C. Perez, J. Snedecor, C. Zlatkov, G. Padmabandu, L. Devesse, S. Radecke, C. L. Holt, S. A. Kumar, B. Budowle, K. M. Stephens, Developmental validation of the ForenSeq Kintelligence kit, MiSeq FGx sequencing system and ForenSeq Universal Analysis Software, Forensic Sci Int Genet, 71(2024), 103055, https://doi.org/10.1016/j.fsigen.2024.103055.

[2] ForenSeq Kintelligence HT Kit Handbook (Rev 03/2024 HB3549-001) https://www.qiagen.com/de-us/products/human-id-and-forensics/nextgeneration-sequencing/forenseq-kintelligence-ht-kit?catno=V16000190 (accessed October 7, 2024).

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 357-364
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Impact of additional STR loci and maternal inclusion on the accuracy of duo and trio paternity tests: A comparative study

Hannah Fontanil[1], Tasnim Munshi[1], Thomas Haizel[2], Sasitaran Iyavoo[1,2*]

[1] *University of Lincoln, Lincoln, United Kingdom.*
[2] *AttoLife Limited, Norwich, United Kingdom.*

** Corresponding author at: Corresponding author at: School of Natural Sciences, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, United Kingdom, siyavoo@lincoln.ac.uk.*

Abstract

Paternity testing is a critical tool in forensic science, often utilised to resolve disputes, establish familial relationships in criminal investigations, and provide evidence in cases of inheritance, custody, or immigration. This study evaluates the impact of using additional STR markers and the inclusion of maternal DNA on the accuracy of paternity testing. The Verifiler™ and Identifiler™ kits were compared in duo (child and putative father) and trio (child, mother, and putative father) paternity tests. Findings indicate that the use of additional markers and the inclusion of maternal DNA significantly enhance the accuracy of paternity testing, reducing the likelihood of false inclusions and exclusions, particularly in duo scenarios. Results show that duo tests, which omit maternal DNA, are more susceptible to errors, especially when fewer markers are analysed. Including the mother and utilising a comprehensive marker set, such as those provided by Verifiler™, improves the reliability of paternity outcomes and minimises the risk of incorrect conclusions. This study recommends duo testing only when absolutely necessary, with the highest possible number of markers employed to ensure reliable results. These findings contribute to the ongoing improvement of paternity testing methodologies, particularly for complex relationships and high-stakes legal cases.

Keywords

Paternity testing, STR markers, Verifiler™, Identifiler™, forensic genetics

1. Introduction

Paternity testing is crucial in forensic applications, resolving legal disputes, identifying familial relationships in criminal investigations, and providing evidence in inheritance, custody, or immigration cases [1]. Early paternity testing used ABO blood group variations, polymorphic proteins, and enzymes, followed by advances with human lymphocyte antigens (HLA). The discovery of hypervariable short tandem repeats (STRs) revolutionised genetic analysis accuracy, a capability that has grown with expanding multiplex PCR kits [2].

DNA paternity testing ideally analyses the mother's and father's alleles at specific STR regions to determine the parental origin of a child's alleles, based on Mendel's laws of inheritance. Duo paternity tests involve only the child and putative father, whereas trio tests include the mother's DNA. With increased demand for duo (motherless) tests, reduced specificity arises since maternal alleles cannot be confirmed without her DNA [3]. Mutations further complicate the interpretation, increasing the likelihood of errors.

Previous studies have explored false inclusions in duo tests [4-5], but false exclusions have received less attention. Given the significant impact of paternity testing, such as in custody cases, false inclusions can be catastrophic. While research has examined the effects of using fewer markers in paternity testing [6-7], this study aims to evaluate the impact of additional STR markers and maternal DNA inclusion on the accuracy of paternity testing. Specifically, it compares the Verifiler™ and Identifiler™ kits in duo and trio tests, assessing differences in Combined Paternity Index (CPI) values, false inclusions, and false exclusions.

2. Material studied, methods, techniques

2.1 DNA profiling

Buccal swabs were collected from 213 cases (156 paternity inclusions and 57 paternity exclusions) previously identified through trio paternity testing. DNA was extracted from all buccal swabs using the Prep-n-Go™ Buffer (Applied Biosystems™) [8]. STR profiling was conducted using the validated reduced-volume PCR method with both the AmpFℓSTR™ Identifiler™ PCR Amplification Kit (Applied Biosystems™), which targets 15 STR loci [9], and the Verifiler™ Express PCR Amplification Kit (Applied Biosystems™), which targets 23 STR loci [10]. Capillary electrophoresis was performed on the ABI Prism® 3500xL Genetic Analyzer

(Applied Biosystems™) according to the manufacturer's instructions, and profile confirmation was conducted using GeneMapper® ID-X v1.6 software (Applied Biosystems™).

2.2 Statistical analysis

The CPI was calculated for both duo and trio paternity scenarios using both kits, with calculations performed on Converge™ Software (Applied Biosystems™), with an internal inclusion acceptance threshold of 10,000 CPI for Verifiler™, and 1,000 CPI for Identifiler™. Significant differences were assessed using a paired t-test, with a p-value threshold set at 0.05.

3. Results

3.1 Inclusion cases

As expected, a positive correlation was identified between the number of loci tested and the CPI in both duo and trio paternity inclusion tests (Figure 1). The paired t-test revealed no significant difference (p-value = 0.067) between the trio paternity tests carried out with both kits. However, a notable difference (p-value = 0.031) was found between the duo paternity tests when additional loci were analysed.

Figure 1: A violin plot presenting the distribution of CPIs in duo and trio tests using both Verifiler™ and Identifiler™ kits.

Using the Verifiler™ kit, the number of cases with CPIs below the acceptance threshold increased from zero to two when the mother was not included. Both of these cases involved mutations, as discussed below. The Identifiler™ kit showed a more pronounced effect, with one case below the threshold in trio tests and a significant rise to eight cases below the threshold in duo tests. Of these, the trio case involved a mutation, whereas only three of the eight duo cases had mutations.

3.2 Exclusion cases

In the exclusion cases, a notable difference was observed between the two kits in duo paternity tests (Figure 2). One of the cases presented only one mismatch with Identifiler™ in a duo test, but this increased to three mismatches with Verifiler™. By including the mother's DNA in a trio test for this case, three mismatches were found with Identifiler™, and seven mismatches with Verifiler™. This demonstrates the importance of including the mother in paternity tests and also the value of additional markers in confidently excluding a putative father. Additionally, when comparing duo paternity tests using Verifiler™ to those using Identifiler™, the average reduction in the number of incompatible loci was 4.4 loci.

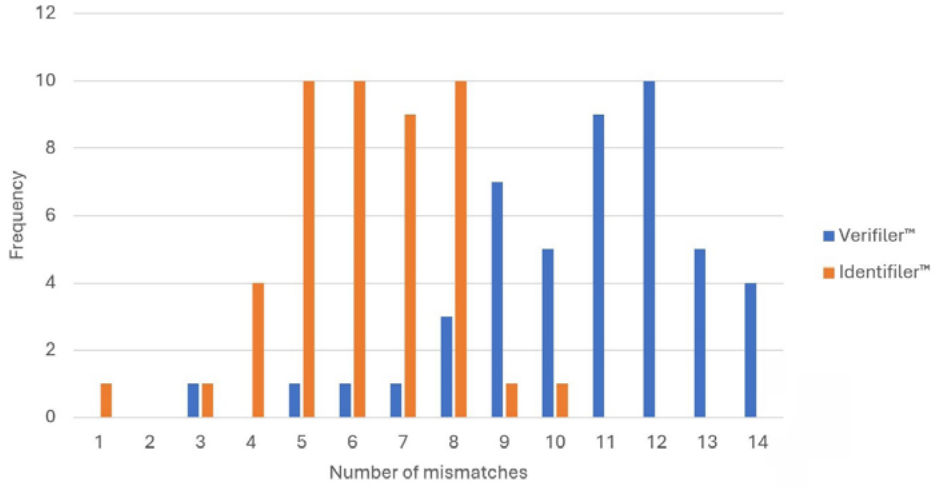


Figure 2: A bar chart displaying the frequency of mismatches in excluded cases in duo tests using both Verifiler™ and Identifiler™ kits.

3.3 Mutations

In paternity testing using STR markers, a mutation refers to a genetic change where the number of repeat units at an STR locus differs between a child and their biological parent. It is considered good practice to assign a mutation when both parents are tested, as testing only one parent could lead to misidentifying a mismatch as an inconsistent allele rather than a mutation, especially in cases involving closely related parents or when a close relative of the putative parent is the biological parent [11]. This can be clarified by including both parents or adding more markers.

In this study, nine mutations were identified: five from the father, two from the mother, and two of undetermined origin. All cases involving paternal mutations produced the lowest CPIs in all types of paternity tests. Considering mismatches as mutations, duo tests using Verifiler™ resulted in CPIs below the inclusion threshold of 10,000 CPI in two cases (392.23 with a one-step deletion at D18S51, and 5,115.94 with a one-step insertion at FGA). This effect was amplified with fewer markers analysed using Identifiler™, where three duo tests fell below the 1,000 CPI threshold, with CPIs of 1.14, 20.80, and 20.21 (the latter involving a one-step mutation of unknown origin at vWA).

Discussion

The significant difference between the duo tests using Verifiler™ and Identifiler™ kits highlights the impact of additional markers on the CPI of these tests. When fewer STRs are analysed, there is a reduction in incompatible loci from excluded cases, with an average reduction of 4.4 loci in this study. Without the mother's DNA, this reduction increases the possibility of wrongful inclusion and exclusion, and this risk is further increased when fewer loci are analysed. This could also be an issue when the mother and father of a child are related, which is common in South Asia through consanguineous marriages between cousins [12]. Such situations could cause issues in immigration cases, where, without the inclusion of the mother and with fewer STRs analysed, there could be a possibility of false inclusion of the putative father, even if he is related to the child but is not the biological father [13]. As shown in this study, adding extra markers to the testing reduces the likelihood of such errors.

Furthermore, the number of cases dropping below each kit's threshold provides valuable insight into the probability of a case being reported as inconclusive

or wrongfully excluded. The fact that the percentage of cases below the threshold in duos decreased from 5.13% with Identifiler™ to 1.28% with Verifiler™ highlights the advantage of additional markers. While not as pronounced, this effect is also mirrored in the trios, with no cases below the threshold with Verifiler™, while 0.64% of cases below the threshold with Identifiler™. In both cases, the increase in cases below the threshold when the mother was not included in the test reveals a key downside of duo paternity tests. While the difference is not drastic between the Verifiler™ trios and duos or between the Verifiler™ trios and the Identifiler™ trios, it remains true that only the Verifiler™ trios experienced no cases dropping below the assigned threshold.

Additionally, when Identifiler™ was used in a duo test on a previously excluded case, only one incompatible locus was identified, increasing the risk of incorrectly attributing this incompatibility to a mutation. The additional markers provided by the Verifiler™ kit increased the number of incompatible loci in this case to three, a number that would prompt further investigation, including statistical evaluation, the addition of more markers, or including the mother in the test, before making a decision on inclusion or exclusion.

While all the data in this study suggest that using more markers and including both parents in the test produces the ideal results, compromising on just one of these variables, if necessary, would not significantly increase the likelihood of a false result. However, as supported by Zhang et al. [14] and Gao et al. [15], using fewer markers and excluding the mother from the test is not recommended.

Conclusion

This study highlights the impact of additional STR markers and maternal DNA inclusion in paternity testing, emphasising the importance of comprehensive approaches for accuracy. Comparing Verifiler™ and Identifiler™ kits shows that extra markers significantly reduce false inclusions or exclusions, particularly in duo tests, which are more error-prone without maternal DNA.

Including both parents and using comprehensive marker sets, like Verifiler™, enhances confidence by reducing ambiguity and minimising errors. This is crucial in cases involving related parents or complex immigration scenarios, where mistakes can have serious consequences.

The study supports duo testing only when necessary, recommending the use of as many markers as possible for reliability. Further research into marker

count, family relationships, and mutation rates will improve paternity testing and mitigate risks.

Conflict of interest statement

All claims in this article are solely those of the authors and do not necessarily reflect their affiliated company or organisation.

References

[1] P.M. Schneider, Scientific standards for studies in forensic genetics, Forensic Sci. Int. 165 (2007) 238-243.

[2] A. Carracedo, Forensic genetics: history. Encyc. Forensic Sci. (2013) 206–210.

[3] M.E. García-Aceves, D.G. Jasso-Razo, X.X. Díaz-Navarro, et al., A posteriori parameters from paternity tests of a Mexican laboratory with the powerplex fusion system, Leg. Med. 64 (2023) 102296.

[4] A.A. De Kock, J.J. Kloppers. The impact of motherless paternity testing in a South African population. Afr. Health Sci. 21(1) (2021) 379–384.

[5] V.R.C. Aguiar, A.M. de Castro, L.M. Pinto, et al., Assessing false paternity risk in simulated motherless cases from more than 20 000 real exclusion trios, Transfusion 61 (2021) 678–681.

[6] K. Babol-Pokora, R. Jacewicz, Pepinski, et al., Identifiler™ system as an inadequate tool for judging motherless paternity cases, Int. Congr. Ser. 1288 (2006) 462–464.

[7] P. Shrivastava, R.K. Kumawat, S. Dixit, Statistical Evaluation of Paternity Using Most Recent STR-Based Multiplex Systems, Proc. Natl. Acad. Sci. India Sect. B. Biol. Sci. 91(3) (2021) 575–580.

[8] S. Iyavoo, S. Knights, M. Mavrommatis, et al., Implementation of Prep-n-Go™ Buffer for DNA extraction from buccal swabs. Forensic Sci. Int. Genet. Suppl. Ser. 7 (2019) 211-213.

[9] S. Iyavoo, S. Cummings, A. Wolejko, et al., Reduced volume PCR amplification using AmpFℓSTR® Identifiler® kit, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e398-e399.

[10] J. Perry, T. Munshi, T. Haizel, et al., Validation of reduced volume VeriFiler™ Express PCR Amplification Kit for buccal swab samples extracted using Prep-n-Go™ Buffer, J. Forensic Sci. 67 (2022) 1971–1978.

[11] C. Børsting, N. Morling, Reinvestigations of six unusual paternity cases by typing of autosomal single-nucleotide polymorphisms, Transfusion 52 (2012) 425–430.

[12] F.Z.A. Khan, S.B. Mazhar, Current trends of consanguineous marriages and its association with socio-demographic variables in Pakistan, Int. J. Reprod. Contracept. Obstet. Gynecol. 7(5) (2018) 1699-1705.

[13] S. Iyavoo, M. Mavrommatis, S. Knights, et al., Heterozygosity reduction in children whose parents are closely related, Forensic Sci. Int. Genet. Suppl. Ser. 8 (2022) 71-73.

[14] M.X. Zhang, H.M. Gao, S.Y. Han, et al., Risk analysis of duo parentage testing with limited STR loci, Genet. Mol. Res. 13(1) (2014) 1179–1186.

[15] H. Gao, C. Wang, R. Zhang, et al., Application of CPI cutoff value based on parentage testing of duos and trios typed by four autosomal kits, PLoS ONE 14(11) (2019) e0225174.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 365-372
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Exploring the long-term DNA stability on buccal 4N6FLOQSwabs® in simulated paternity cases using a traditional and an alternative forensic workflow

Michele Rosso[1*], Anna Franzoni[1]

[1] *Copan Italia S.p.A., Brescia, Italy.*

** Corresponding author at: michele.rosso@copangroup.com*

Abstract
This study evaluated the long-term DNA stability on buccal 4N6FLOQSwabs® stored at room temperature for over five years in six simulated paternity cases. A traditional workflow and the microFLOQ® direct amplification method were assessed, each allowing to effectively establish the eventual paternity attribution.

Keywords
Stability, paternity, 4N6FLOQSwabs®, microFLOQ®, subsampling.

Introduction

Paternity testing is conducted to determine whether an individual is the biological father or mother of a child. DNA parentage testing is the most reliable method for establishing or excluding biological parentage [1]. This involves the comparison of STR profiles, considering that biological parents share half of their alleles with the child. A likelihood ratio (LR) exceeding the nominal threshold of 10,000 is conventionally required to support the genetic evidence and establish paternity with a high degree of biostatistical certainty [2].

A common approach to postnatal paternity testing involves the analysis of genetic information obtained from buccal swabs [3]. 4N6FLOQSwabs® (Copan Italia) (4N6FS) are flocked nylon swabs, certified free of amplifiable human DNA, that are widely used for crime scene [4,5] and buccal sample collection [3].

This study aimed to explore the long-term DNA stability on buccal 4N6FS with active drying system, stored at RT for over 5 years, in six simulated paternity cases. Verifying the stability of DNA on the device over years addresses contexts where delays in testing can occur or when extensive legal review and validation are necessary.

In addition to the standard analysis of 4N6FS, which includes DNA extraction, quantification, and STR profiling (standard workflow), an alternative workflow utilizing microFLOQ® direct (Copan Italia) (MF) was evaluated (direct workflow). MF allows for direct amplification because its tip is treated with a lysing agent eliminating the need for DNA extraction and quantification [6,7] and enabling the collection of a small portion of the sample, thus preserving it for additional testing [6,8]. By using MF, 4N6FS were subsampled to expedite the analytical process.

Material studied, methods, techniques

Sample collection and preparation

Starting from 2018, six simulated paternity cases have been set-up collecting two buccal 4N6FS from each individual of interest. One 4N6FS was analyzed immediately after collection (T0), while the second following long-term storage at RT, after the time-to-analysis (TTA) interval detailed below:

— Case 1 (T0:07/2018–TTA:5 years and 8 months): Daughter, Mother;

— Case 2 (T0:11/2019–TTA:4 years and 4 months): Son, Mother, Father;

— Case 3 (T0:01/2020–TTA:4 years and 2 months): Son, Mother, Father;

— Case 4 (T0:09/2020–TTA:3 years and 6 months): Son, Mother, Father;

— Case 5 (T0:07/2021–TTA:2 years and 8 months): Daughter, Father;

— Case 6 (T0:10/2021–TTA:2 years and 4 months): Son, Father.

At TTA, before the analysis, each 4N6FS was subsampled using a MF pre-wetted with 1µl of PCR grade water.

Standard workflow

The standard (Std) workflow included automated extraction of DNA using QIAcube (Qiagen) from each 4N6FS, utilizing the QIAamp® DNA Investigator Kit (Qiagen) according to the manufacturer's instructions, with the following modifications: after 4N6FS lysis (Isolation of DNA from surface and buccal swabs,

Part A, V2, May 2008), the swab tip was placed in a NAO® Basket (Copan Italia) and centrifuged at maximum speed for 1 minute. The residual volume was added to the lysate for the subsequent purification protocol (Isolation of DNA from surface and buccal swabs, Part B, V2, May 2008), with a final elution volume of 100 µl.

DNA quantification was carried out using real-time PCR with the Quantifiler® Trio DNA Quantification kit on the AB7500 Real-Time PCR System (Thermo Fisher Scientific). The reactions and thermal cycling were set up in the HID Real-Time PCR Analysis software v1.2, following the manufacturer's instructions.

The optimal 1ng DNA input was amplified using the GlobalFiler™ PCR Amplification Kit on Veriti™ 96-well, as per manufacturer's instructions for 29 PCR cycles (Thermo Fisher Scientific). Capillary electrophoresis for STR typing was run on AB3500 Genetic Analyzer (Thermo Fisher Scientific), based on the kit's user manual indications, utilizing an internally validated injection setting of 1.2kV for 15 seconds. STR profiles were analyzed with GeneMapper ID-X v1.6 software (Thermo Fisher Scientific), applying internally validated analytical and stochastic thresholds at 135 RFU and 675 RFU, respectively.

Direct workflow

Each MF tip was broken into a PCR reaction tube for direct amplification using the GlobalFiler™ Express PCR Amplification Kit at 28 PCR cycles on a Veriti™ 96-well thermal cycler (ThermoFisher Scientific). Capillary electrophoresis for STR typing was run on AB3500 Genetic Analyzer (Thermo Fisher Scientific), following the kit's user manual instructions, with an internally validated injection setting of 1.2 kV for 8 seconds. *STR profiles were analyzed with GeneMapper ID-X v1.6 software (Thermo Fisher Scientific), applying internally validated analytical and stochastic thresholds at 235 RFU and 700 RFU, respectively.*

STR profiling and statistical analyses

Obtained STR profiles were assessed based on:

— *Average peak heights (PH) of all allelic calls, expressed in RFU;*

— *Average peak height ratio (PHR),* expressed as a percentage and calculated for each heterozygous locus as the ratio between the lower to the higher peak height;

— Average intra-color balance (ICB), calculated as the ratio between the minimum PH value in a fluorescent dye channel to the average PH of all alleles in the same fluorescent dye channel.

Statistical analysis was conducted using Minitab v21 software, utilizing descriptive statistics and a two-sided Student's t-test to analyze the grouped variables of interest.

Results

This study analyzed 15 buccal 4N6FS with Std and direct workflows.

Results of Quantifiler® Trio DNA Quantification kit on 4N6FS extracted at the TTA showed variability in the quantity of DNA, with the lowest amount detected in sample from the son of case 2 (*Table 1*).

Table 1. Average quantity of DNA (ng/µl) and degradation index ± Standard Deviation of 4N6FS extracted at TTA.

Case #	Stability	Donor	T.Large autosomal (ng/µl)	T.Small autosomal (ng/µl)	T.Y (ng/µl)	Degradation Index
1	5 years + 8 months	Daughter	12,40 ± 0,07	16,74 ± 0,05	n.a.	1,35 ± 0,01
		Mother	1,00 ± 0,02	1,13 ± 0,17	n.a.	1,13 ± 0,15
2	4 years + 4 months	Father	39,25 ± 0,23	42,47 ± 1,37	36,97 ± 1,56	1,08 ± 0,03
		Son	0,31 ± 0,01	0,40 ± 0,02	0,41 ± 0,02	1,29 ± 0,03
		Mother	0,73 ± 0,06	0,59 ± 0,20	n.a.	0,80 ± 0,21
3	4 years + 2 months	Father	1,80 ± 0,31	4,19 ± 3,69	2,46 ± 0,71	2,18 ± 1,48
		Son	6,80 ± 0,12	10,47 ± 0,88	10,29 ± 0,60	1,54 ± 0,16
		Mother	17,44 ± 0,76	24,63 ± 0,06	n.a.	1,41 ± 0,07
4	3 years + 6 months	Father	3,01 ± 0,00	2,95 ± 0,09	2,57 ± 0,14	0,98 ± 0,03
		Son	2,35 ± 0,04	2,98 ± 0,01	2,92 ± 0,07	1,27 ± 0,02
		Mother	3,89 ± 0,67	3,70 ± 1,16	n.a.	0,94 ± 0,14
5	2 years + 8 months	Father	4,24 ± 0,06	5,35 ± 0,58	5,45 ± 0,12	1,26± 0,12
		Daughter	0,82 ± 0,02	1,15 ± 0,12	n.a.	1,40± 0,13
6	2 years + 4 months	Father	3,30 ± 0,03	3,34 ± 0,02	3,09 ± 0,04	1,01 ± 0,02
		Son	3,43 ± 0,03	3,12 ± 0,14	2,92 ± 0,04	0,91 ± 0,05

Complete and concordant STR profiles from the same individuals were obtained at TTA versus T0 analysis, using both standard and direct workflows, except for two MF profiles: one showed a missing locus, and the other displayed allele drop-out at two different markers (*Table 2*).

Table 2. Results of STR profiling with number of comparable markers and alleles obtained with Std and direct workflows versus T0 analysis.

CASE #	Buccal swab sample		T0 analysis assay (marker n. /allele n.)	Std workflow - GlobalFiler					MF workflow - GlobalFiler Express				
	Donor	Stability		Full profile	n. of comparable markers	Marker match (%)	n. of comparable alleles	Allele match (%)	Full profile	n. of comparable markers	Marker match (%)	n. of comparable alleles	Allele match (%)
1	Daugther	5 years + 8 months	Global Filer (22/44)	YES	22	22 (100%)	44	44 (100%)	YES	22	22 (100%)	44	44 (100%)
	Mother		Global Filer (22/44)	YES	22	22 (100%)	44	44 (100%)	YES	22	22 (100%)	44	44 (100%)
2	Father	4 years + 4 months	Verifiler Express (25/49)	YES	22	22 (100%)	43	43 (100%)	YES	22	22 (100%)	43	43 (100%)
	Son		Verifiler Express (25/49)	YES	22	22 (100%)	43	43 (100%)	NO[b]	22	21 (95,4%)	43	41 (95,3%)
	Mother		Verifiler Express (24/48)	YES	21	21 (100%)	42	42 (100%)	YES	21	21 (100%)	42	42 (100%)
3	Father	4 years + 2 months	GlobalFiler (24/46)	YES	24	24 (100%)	46	46 (100%)	YES	24	24 (100%)	46	46 (100%)
	Son		GlobalFiler (24/46)	YES	24	24 (100%)	46	46 (100%)	YES	24	24 (100%)	46	46 (100%)
	Mother		Global Filer (22/44)	YES	22	22 (100%)	44	44 (100%)	YES	22	22 (100%)	44	44 (100%)
4	Father	3 years + 6 months	GlobalFiler (24/46)	YES	24	24 (100%)	46	46 (100%)	YES	24	24 (100%)	46	46 (100%)
	Son		GlobalFiler (24/46)	YES	24	24 (100%)	46	46 (100%)	YES	24	24 (100%)	46	46 (100%)
	Mother		Global Filer (22/44)	YES	22	22 (100%)	44	44 (100%)	YES	22	22 (100%)	44	44 (100%)
5	Father	2 years + 8 months	GlobalFiler Express (24/46)	YES	24	24 (100%)	46	46 (100%)	YES	24	24 (100%)	46	46 (100%)
	Daughter		GlobalFiler Express (22/44)	YES	22	22 (100%)	44	44 (100%)	NO[b]	22	20 (90,9%)	44	42 (95,4%)
6	Father	2 years + 4 months	GlobalFiler (24/46)	YES	24	24 (100%)	46	46 (100%)	YES	24	24 (100%)	46	46 (100%)
	Son		GlobalFiler (24/46)	YES	24	24 (100%)	46	46 (100%)	YES	24	24 (100%)	46	46 (100%)

[a] = TPOX locus drop-out;

[b] = 2 allelic drop-out in different loci: missing allele "7" in TH01 and allele "12" in D7S820.

Variability was detected in average PH (Std workflow: 6567 RFU; direct workflow: 4275 RFU, considering only complete MF profiles), average heterozygous PHR (Std workflow: 83,57; direct workflow: 81,75 considering only complete MF profiles), and average ICB (Std workflow: B:0,56; G:0,49; Y:0,67; R:0,61; P:0,61; direct workflow: B:0,51; G:0,45; Y:0,58; R:0,44; P:0,38 considering only complete MF profiles) (*Figure 1*).

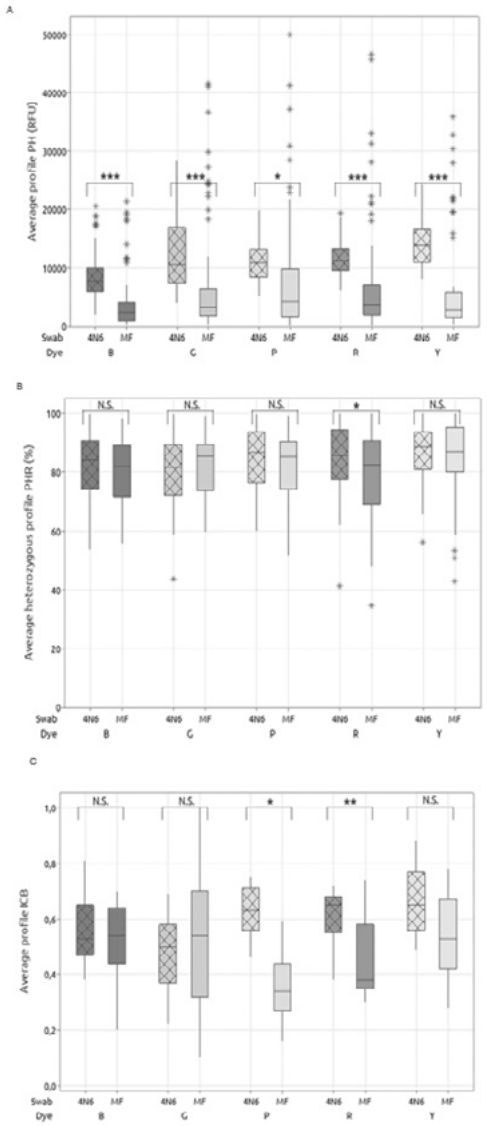


Fig. 1. Boxplots comparing STR profiles obtained from 4N6FS samples (crossed boxes) and their respective MF subsamples (empty boxes):

A: Average PH (RFU) grouped by fluorescent dye and swabbing method;

B: Average heterozygous PHR (%) grouped by fluorescent dye and swabbing method;

C: Average ICB grouped by fluorescent dye and swabbing method.

(NS) when p-value is >0,05; (*) when p-value is <0,05; (**) when p-value is <0,01; (***) when p-value is <0,001.

Discussion

Regarding the Std workflow, despite the variability in the amount of DNA extracted from buccal 4N6FS, which sometimes resulted in low yields, the optimal input for the STR typing PCR was always achieved. Even when degradation index exceeded 1, complete profiles were still generated. The comparable alleles obtained from the STR profiling kits used at T0 and TTA were consistent for all individuals.

Although there was variability in PH, PHR, and ICB, STR profiles were complete for all the individuals when using the Std workflow, while incomplete profiles from MF in case 2 and case 5 were likely due to low DNA yield of initial buccal 4N6FS samples (*Table 1*). The comparable alleles obtained from the STR profiling kits used at T0 and TTA were consistent for all individuals also when using the direct workflow. The cumulative PH averages obtained using the direct workflow were significantly lower than the respective values generated with the standard workflow across all the fluorescence dye channels (*Figure 1A*). Instead, the cumulative PHR and ICB averages obtained with both workflows were comparable, except for significant differences in red and purple fluorescence dye channels, with worse results obtained with MF (*Figure 1B, 1C*). The lower quality results generated by MF can be attributed to the direct amplification, along with the low amount of subsampled DNA after long-term storage. Despite this, the STR profiles obtained at TTA using both workflows allowed for the same conclusions regarding the kinship hypotheses defined at T0 for all cases. Indeed, although the direct workflow resulted in the exclusion of one locus from the profile of case 2 and two loci in case 5, paternity attribution was still established with the LR threshold above 10,000.

Conclusion

Results demonstrated that after prolonged RT storage for over 5 years, buccal 4N6FS allowed for successful generation of complete STR profiles.

The direct workflow including MF has the potential to expedite the paternity testing process while preserving the original sample for further analysis.

Acknowledgments

No acknowledgments to declare.

Conflict of interest statement

At the time of manuscript submission, the authors are employees of Copan Italia S.p.A., the manufacturer of 4N6FLOQSwabs® and microFLOQ® direct. Opinions, findings, and conclusions expressed in this publication are those of the authors and do not necessarily reflect those of Copan Italia S.p.A.

References

[1] M. Patidar, S. Agrawal, F. Parveen et al. Molecular insights of saliva in solving paternity dispute. J Forensic Dent Sci. (2015) 7(1):76-9.

[2] S. Willis, L. Kenna, S. Dermott et al. ENFSI guideline for Evaluative reporting in forensic science. Strengthening the Evaluation of Forensic Results across Europe. ENFSI (2022).

[3] J.C.W. Tam, Y.M. Chan, S.Y. Tsang et al. Noninvasive prenatal paternity testing by means of SNP-based targeted sequencing. Prenat Diagn. (2020) 40(4):497-506.

[4] L. Jansson, C. Forsberg, Y. Akel et al. Factors affecting DNA recovery from cartridge cases. Forensic Sci Int Genet. (2020) 48:102343.

[5] B. Ferreira-Silva, M.J. Porto, T. Magalhães et al. Contribution to the Development of Guidelines in the Analysis of Biological Evidence in Sexual Assault Investigations. J Forensic Sci. (2019) 64(2):534-538.

[6] C. Loockerman, B. Miller, R. Ramsey et al. Collection and storage of DVI samples with microFLOQ® Direct swabs for direct amplification. Forensic Sci Int Genet. (2021) 55:102588.

[7] A.J. Sherier, R.e. Kieser, N.M.M. Novroski et al. Copan microFLOQ® Direct Swab collection of bloodstains, saliva, and semen on cotton cloth. Int J Legal Med. (2020) 134(1):45-54.

[8 R. Shadoff, J. McGregor,K. Britton et al. Direct Amplification of Blood, Saliva, and Semen Subsamples Using Copan MicroFLOQ® Direct Swabs. Forensic Genom. 4(3).

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 373-378
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Forensic complex kinship inferences with a combination of STRs and MHs accounting for linkage

Shanshan Wei[1,a], Jiaming Xue[1,a], Qiushuo Wu[1], Mengyu Tan[1], Yazi Zheng[1], Guihong Liu[1], Ranran Zhang[1], Yuanyuan Xiao[1], Haoyan Jiang[1], Meili Lv[2], Shengqiu Qu[1], Lin Zhang[1*], Weibo Liang[1*]

[1] *Department Of Forensic Genetics, West China School Of Basic Medical Sciences And Forensic Medicine, Sichuan University, Chengdu, Sichuan 610041, China.*

[2] *Department Of Immunology, West China School Of Basic Medical Sciences And Forensic Medicine, Sichuan University, Chengdu, Sichuan 610041, China.*

[a] *These authors contributed equally to this work.*

** Corresponding author at: Department of Forensic Genetics, West China School of Basic Medical Sciences and Forensic Medicine, Sichuan University, Chengdu, Sichuan 610041, China. zhanglin@scu.edu.cn (L. Zhang), liangweibo@scu.edu.cn (W. Liang).*

Abstract

As prominent genetic markers, microhaplotypes (MHs) offer several advantages over short tandem repeats (STRs), such as lower mutation, non-preferential amplification, no stutter, and short amplicon fragments. MHs have been utilized in individual identification, DNA mixture analysis, complex kinship inference, ancestral information provision, and non-invasive prenatal relationship testing. Kinship identification is essential for investigating crimes, finding lost individuals, recognizing victims of disasters, and resolving inheritance conflicts. In the absence of other relatives, determining the kinship relationships between distant relatives can be quite challenging. Therefore, increasing the number of genetic markers is essential for improving the system power of kinship inferences. The influence of linkage between loci should be considered in kinship analysis. In this study, four panels (STR + Panel A, STR + Panel B, Panel A & B, and STR + Panel A & B) were constructed. Then, 198 samples from unrelated families to evaluate the efficiency of various kinship relationship types were collected. As a result, the samples were all successfully genotyped. STR + Panel A & B demonstrated the highest system power among the four panels. For STR + Panel A & B, when |t| was 4 or 5, the sensitivity, specificity, and effectiveness clocked up 1 with no error in full-siblings (FS). The sensitivity, specificity, and effectiveness were above 0.96 with no error in 2nd-degree relatives of half-siblings (HS), uncle/aunt-nephew/niece (UN),

and grandparents-grandson/granddaughter (GS). However, the sensitivity, specificity, and effectiveness of 3rd-degree relatives were inferior compared with FS and 2nd-degree relatives.

Keywords
Kinship identification, linkage, microhaplotypes, short tandem repeats.

Introduction

Microhaplotypes (MHs), recognized as powerful genetic markers, are defined as a combination of alleles that contains two or more spaced single nucleotide polymorphisms (SNPs) within a short distance and has a fragment length of less than 300 bp, which initially propounded by Prof. Kidd in 2013 [1]. MHs have the advantages of short tandem repeats (STRs) and SNPs, such as a lower mutation rate, non- preferential amplification, no stutter products, short amplicon fragments, and higher levels of polymorphism, which effectively perform individual identification, analyze DNA mixtures, identify complex kinship inferences, provide ancestral information, and conduct non-invasive prenatal relationship testing [2]. Kinship identification, as an important component of forensic genetics, confirms the existence of genetic relationships among different individuals by detecting genetic markers. Moreover, kinship identification plays a crucial role in criminal investigations, locating missing persons, identifying disaster victims, and resolving inheritance disputes [3-5]. Linkage refers to the tendency of genes located on the same chromosome to be co-inherited, a phenomenon that arises from recombination events within a pedigree. In this study, we conducted four panels: STR + Panel A, STR + Panel B, Panel A & B, and STR + Panel A & B, and gathered 198 samples from unrelated families to evaluate the efficiency of various real and simulated kinship relationship pairs with fully accounting for linkage.

Materials and methods

In this study, a total of 198 samples from 14 unrelated extended families of Southwest Chinese Han individuals were collected and sequenced for both MH panels (53 loci of Panel A and 67 loci of Panel B), with 19 STRs genotyped by Capillary electrophoresis. The average effective numbers of alleles (A_e) of Panel A and Panel B were 7.43 and 5.34, respectively. No linkage disequilibrium (LD) was found among those markers. Amongst these samples, 133 individuals were collected and sequenced for

MH panels based on next-generation sequencing (NGS) results from previous studies [6]. The samples contained 80 full-siblings (FS); 128 uncle/aunt-nephew/niece (UN), 89 grandparents-grandson/granddaughter (GS), and 8 half-siblings (HS); 131 3rd-degree, 70 4th-degree, 39 5th-degree, and 25,137 unrelated duos. Simulations were conducted for both related and unrelated cases of each kinship hypothesis of 1000 times. Moreover, Merlin v1.1.2 software was used for pairwise kinship analysis with linked markers across four panels (STR + Panel A, STR + Panel B, Panel A & B, and STR + Panel A & B) based on simulated and real family data.

Results

All samples were successfully genotyped. The discrimination power of the kinship inferences of the four panels (STR + Panel A, STR + Panel B, Panel A & B, and STR + Panel A & B) is shown as Log10 of the likelihood ratio (Log10LR). STR + Panel A & B (Table 1) exhibited the highest system power among four panels, followed by Panel A & B, STR + Panel A, and STR + Panel B.

In FS, the average values of log10LR respectively were 27.16 ± 0.98, 27.20 ± 0.94, 43.81 ± 1.50, and 50.19 ± 1.70 for STR + Panel A, STR + Panel B, Panel A & B and STR + Panel A & B. And when |t| was equal to 4 or 5, the sensitivity clocked up 1, while both specificity and effectiveness exceeded 0.999 with no error for all panels.

In 2nd-degree relatives, the average log10LR values were 8.13 ± 0.35, 7.88 ± 0.39, and 9.05 ± 1.07 in UN, GS, and HS for STR + Panel A, 7.58 ± 0.30, 7.79 ± 0.40, and 8.21 ± 0.96 for STR + Panel B, 12.58 ± 0.46, 13.57 ± 0.58, and 12.93 ± 1.43 for Panel A & B, and 14.47 ± 0.52, 15.42 ± 0.63, and 15.93 ± 0.63 for STR + Panel A & B, respectively. With a log10LR threshold set at 3, the sensitivity, specificity, and effectiveness were above 0.88 (error rate below 0.0004) for STR + Panel A and STR + Panel B, and above 0.96 (error rate below 0.00004) for Panel A & B and STR + Panel A & B.

In 3rd-degree relatives, the average value of log10LR was 4 ± 0.26 for STR + Panel A & B. For STR + Panel A and STR + Panel B, the sensitivity, specificity, and effectiveness were even below 0.8 (error rate 0.00638) at |t| = 1, and below 0.5 (error rate above 0.0005) at |t| =2. In addition, the specificity and effectiveness were over 0.92 for Panel A & B, while above 0.93 for STR + Panel A & B at |t| = 1 (error rate below 0.002). When |t| was 2, the sensitivity, specificity, and effectiveness were above 0.67 (error rate 0.0002) and 0.72 (error rate 0.00024) for Panel A & B and STR + Panel A & B, respectively.

Table 1. The system power of STR + Panel A & B to discriminate between full-siblings (FS), 2nd-degree relatives (UN, GS, and HS), and 3rd-degree relatives from unrelated in real data.

Panel	Relationship	t1(log10)	t2(log10)	Sensitivity	Specificity	PPV	NPV	Error rate	Inconclusive	Effectiveness
STR + Panel A & B	FS	4	−4	1	1	1	1	0	0	1
		5	−5	1	1	1	1	0	0	1
	GS	1	−1	1	0.9994	1	1	0	0.00059	0.99941
		2	−2	0.98876	0.99753	1	1	0	0.0025	0.9975
		3	−3	0.97753	0.98962	1	1	0	0.01043	0.98957
		4	−4	0.97753	0.96392	1	1	0	0.03603	0.96397
	UN	1	−1	0.99219	0.99976	0.99219	1	0.00004	0.00024	0.99972
		2	−2	0.99219	0.99869	1	1	0	0.00135	0.99865
		3	−3	0.99219	0.99527	1	1	0	0.00475	0.99525
		4	−4	0.98438	0.98417	1	1	0	0.01583	0.98417
	HS	1	−1	1	0.99972	1	1	0	0.00028	0.99972
		2	−2	1	0.99829	1	1	0	0.00171	0.99829
		3	−3	1	0.99443	1	1	0	0.00557	0.99443
		4	−4	1	0.9817	1	1	0	0.01829	0.98171
	3rd-degree	1	−1	0.86598	0.93297	0.67742	0.99983	0.00174	0.06555	0.93271
		2	−2	0.72165	0.76891	0.93333	0.99995	0.00024	0.23104	0.76872

FS: full-sibling; HS: half-sibling; GS: grandparents-grandson/granddaughter; UN uncle/aunt-nephew/niece; PPV: positive predictive value; NPV: negative predictive value.

Discussion

With the relationships among multiple individuals remaining unknown, determining the kinship relationships between distant relatives is quite challenging, as distant relatives do not have same alleles at certain loci, while unrelated individuals may happen to possess the same alleles at certain loci. Therefore, increasing the number of genetic markers is crucial for improving the system power of kinship inferences. In this study, four panels (STR + Panel A, STR + Panel B, Panel A & B, and STR + Panel A & B) were analyzed. The STR + Panel A & B combination exhibited the highest system power among four panels. Furthermore, the effectiveness of the panel declines with an increase in the degree of kinship. These findings indicated the critical importance of increasing the number of loci and selecting appropriate genotype markers for kinship identification. In addition, LD can influence the transmission of genetic information across different loci, which helps explain differences in systemic power in some kinship identification. In summary, the results not only validated the differential effectiveness of various

genotype combinations in kinship inference, but also emphasized the necessity of considering the hierarchy and complexity of the kinship identification in the following research.

Conclusions

In conclusion, we successfully evaluated the system power of four panels (STR + Panel A, STR + Panel B, Panel A & B, and STR + Panel A & B) in kinship inferences. STR + Panel A & B demonstrated the highest system power among four panels with the average log10LR values of 50.19 ± 1.70, 14.47 ± 0.52, 15.42 ± 0.63, 15.93 ± 0.63, and 4 ± 0.26 in FS, UN, GS, HS and 3rd-degree relatives, respectively. This provides a significant support and guidance for further research in kinship identification.

Acknowledgments

We extend our gratitude to all volunteers who contributed their samples for this study.

Conflict of interest statement

The authors declare no competing interests.

References

[1] K.K. Kidd, A.J. Pakstis, W.C. Speed, R. Lagace, J. Chang, S. Wootton, N. Ihuegbu, Microhaplotype loci are a powerful new type of forensic marker, Forensic Sci. Int.: Genet. Suppl. Ser. 4 (2013) e123–e124.

[2] F. Oldoni, K.K. Kidd, D. Podini, Microhaplotypes in forensic genetics, Forensic Sci. Int.: Genet. 38 (2019) 54–69.

[3] S.K. Mo, Z.L. Ren, Y.R. Yang, Y.C. Liu, J.J. Zhang, H.J. Wu, Z. Li, X.C. Bo, S. Q. Wang, J.W. Yan, M. Ni, A 472-SNP panel for pairwise kinship testing of second-degree relatives, Forensic Sci. Int.: Genet. 34 (2018) 178–185.

[4] H.K. Brustad, M. Colucci, M.A. Jobling, N.A. Sheehan, T. Egeland, Strategies for pairwise searches in forensic kinship analysis, Forensic Sci. Int.: Genet. 54 (2021).

[5] Y.M. Zhang, Z.L. Yu, X.T. Mo, X.C. Zhao, W.S. Li, H. Liu, C. Liu, R.G. Wu, H.Y. Sun, Comparative evaluation of autosomal STRs and X-chromosome STRs as a complement of autosomal STRs in kinship testing in Southern Han Chinese, Ann. Hum. Biol. 48 (1) (2021) 66–69.

[6] J.M. Xue, M.Y. Tan, R.R. Zhang, D.Z. Chen, G.H. Liu, Y.Z. Zheng, Q.S. Wu, Y.Y. Xiao, M. Liao, S.Q. Qu, W.B. Liang, Evaluation of microhaplotype panels for complex kinship analysis using massively parallel sequencing, Forensic Sci. Int.: Genet. 65 (2023) 102887.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 379-384
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Kinship testing using X-STR linkage groups on a Tamil pedigree

Sharlize Pedroza-Matute[a,b], Thomas Haizel[b], Sasitaran Iyavoo[a,b*]

[a] *University of Lincoln, Lincoln, United Kingdom.*
[b] *AttoLife Limited, Norwich, United Kingdom.*

** Corresponding author at: School of Natural Sciences, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, United Kingdom, siyavoo@lincoln.ac.uk.*

Abstract

X-chromosome Short Tandem Repeat (X-STR) markers are valuable supplements to autosomal STR testing in complex kinship cases, such as deficiency paternity and incest. The Investigator® Argus X-12 QS Kit enables multiplex typing of twelve X-STRs, clustered into four linkage groups, producing three-allele mini-haplotypes from male samples. This study evaluated the kit for kinship testing in a Tamil pedigree, involving eight males from three sets of full siblings who shared maternal grandparents, resulting in eight full sibling pairs and twenty first cousin pairs. Mothers' samples were included to confirm maternity and full sibling relationships. Buccal swabs were processed using Prep-n-Go™ Buffer and amplified with the Argus X-12 QS Kit. The VeriFiler™ Express PCR Amplification Kit was used to successfully confirm all relationships. The X-STR results showed full male siblings shared one or two linkage groups. Among first cousins, five pairs shared no linkage groups, eight shared one, four shared two, and three shared three. The study demonstrates that first cousins with the same maternal grandparents can share more linkage groups than full siblings. Additionally, two rare events were observed: a recombination within a linkage group and a mutation. This study underscores the need for careful evaluation when using the Argus X-12 QS Kit for pedigree resolution involving first cousins, while recognising the potential for rare events.

Keywords

Investigator® Argus X-12 QS Kit, X-STRs, linkage groups, kinship testing, recombination, mutation.

Introduction

Short Tandem Repeat (STR) markers located on the X chromosome (X-STRs) can supplement other genetic markers in kinship testing and complex forensic cases.

Their utility is particularly evident in deficiency paternity cases, where the putative father is unavailable for testing, and DNA profiles from sisters or stepsisters are sufficient to exclude paternity. X-STR testing is also valuable in incest cases, especially when the potential fathers are related, such as father and son. In these scenarios, the power of X-STRs can be considerably higher than that of autosomal markers [1-2]. This is due to the specific inheritance pattern of the X chromosome: females inherit one X chromosome from each parent, while males inherit only one from their mother. Recombination occurs only in females, meaning that sisters who share the same father inherit the same paternal X chromosome, while the maternally inherited X chromosome results from the recombination of their mother's two X chromosomes.

However, because of the limited length of the X chromosome, the number of genetic markers that are independently transmitted to offspring is restricted. Only four linkage groups have been identified in which recombination is not expected [1,3-4]. The Investigator® Argus X-12 QS Kit, developed by Qiagen, offers multiplex typing of twelve X-STRs clustered into these four linkage groups, producing mini-haplotypes of three alleles from male samples [5]. Research into the use and characteristics of X-STRs is essential to fully support their application in forensic cases. To promote the utilisation of X-STRs, the utility of the Argus X-12 QS Kit for kinship testing was evaluated in a Tamil pedigree.

Material studied, methods, techniques

Study samples

After obtaining informed consent, samples were collected from eight Malaysian Indian (Tamil) male individuals, grouped into three sets of full siblings (Figure 1). The siblings shared the same maternal grandparents, resulting in eight pairs of full siblings and twenty pairs of first cousins. Samples were also collected from their mothers.

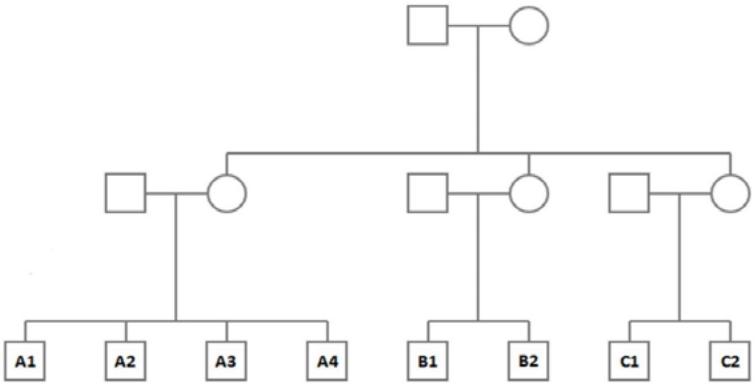


Figure 1: Tamil family pedigree.

DNA extraction and STR typing

Buccal swabs were processed using Prep-n-Go™ Buffer (Applied Biosystems™) [6] and amplified with the Investigator® Argus X-12 QS Kit (Qiagen), following the manufacturer's recommendations. Capillary electrophoresis was performed on an ABI Prism® 3500xL Genetic Analyzer, and DNA profiles were generated using GeneMapper® ID-X v1.6 software.

Confirmation of relationships

Confirmatory relationship tests were conducted using the VeriFiler™ Express PCR Amplification Kit (Applied Biosystems™) [7], applying maternity and full sibling probability thresholds of 99.99% and 95%, respectively.

Results

Maternity and full sibling relationships were all confirmed. Typing of the mothers' profiles revealed two rare events: a recombination between DXS10148 and DXS10135 within linkage group 1, and a one-step repeat-loss mutation at DXS10134. As highlighted in Figure 2, each pair of full siblings shared a minimum of one and a maximum of two linkage groups. Conversely, while five pairs of first cousins shared no linkage groups, eight pairs shared one, four pairs shared two, and three pairs shared three linkage groups.

	A1	A2	A3	A4	B1	B2	C1	C2
A1		1	1	2	3	2	1	3
A2	1		2	2	1	0	1	1
A3	1	2		2	1	1	1	0
A4	2	2	2		2	0	0	1
B1	3	1	1	2		1	0	2
B2	2	0	1	0	1		2	3
C1	1	1	1	0	0	2		2
C2	3	1	0	1	2	3	2	

Figure 2: Number of shared linkage groups between the tested individuals.

Discussion

X-STRs have proven to be informative markers for determining complex relationships, providing stronger statistical support when used in combination with autosomal markers, compared to autosomal testing alone. For instance, in a paternity case involving the mother, daughter, and the mother of the putative father, the Likelihood Ratio (LR) calculated from autosomal markers increased by almost $5x10^8$ when the Argus X-12 X-STRs were added. As a result, the probability of paternity increased from 93% to 99.99999%, significantly enhancing the power of the test [4]. Additionally, X-STR analysis has shown potential in distinguishing certain second-degree relationships, which autosomal markers are unable to do. In attempts to differentiate between grandparental, avuncular, and half-sibling relationships, satisfactory LRs were obtained from the analysis of Argus X-12 markers only in cases where genetic incompatibilities were present. In other cases, low LRs were observed. Based on these findings, caution is recommended when using Argus X-12 markers alone to resolve pedigrees [8].

The results of this study emphasise the need for careful evaluation when using X-STRs and linkage groups for kinship assessment. It was found that first cousins with the same maternal grandparents could share more linkage groups than full siblings. This underscores the fact that analysing X-STR linkage groups alone is not always reliable for inferring biological relationships and should be approached with caution.

Furthermore, two rare events were observed in this study: a mutation and a recombination within a linkage group. A deficiency paternity case involving two

sisters and their mother had previously been reported, in which three rare events occurred [9]. Genetic mapping of X-STR loci has highlighted the presence of non-negligible recombination both within and between linkage groups, emphasising the importance of considering recombination rates when using these markers [10]. Accurate estimation of mutation rates is crucial for reliable interpretation of genetic evidence, as it can determine whether hypotheses are accepted or rejected. However, more research is needed to increase the amount of available data for different populations. While some kinship issues may not require consideration of linkage, in others ignoring it could result in either over- or under-quantification of the genetic evidence. Therefore, further efforts should be made to address these challenges and improve practice [11].

This study was conducted using a Tamil pedigree from the Malaysian Indian minority. A recent study examined X-STR frequencies in Malaysian Malay and Malaysian Chinese individuals, highlighting the importance of evaluating population data for ethnic minorities and admixed populations. However, the Indian minority, which represents approximately 6.6% of the Malaysian population, was not included in that study [12].

To ensure continued progress and support for the use of Argus X-12 kit markers, it is essential to evaluate population data, including mutation and recombination rates, to guarantee fair representation of all ethnicities.

Conclusion

This study highlights the importance of careful evaluation when using the Argus X-12 QS Kit for pedigree resolution. While X-STR testing is a valuable tool in various forensic scenarios, its application in kinship testing involving first cousins with the same maternal grandparents should account for the possibility that cousins, by chance, may share more linkage groups than full brothers. The resolution of relationship cases should also consider the potential co-occurrence of rare events, such as recombination between closely located markers within the same linkage group.

Further studies should continue to investigate the characteristics of X-STRs, including mutation and recombination rates, to support their use in routine forensic cases. Larger and more diverse populations should also be explored to enhance databases, ensuring that data is available for a wide range of case scenarios. This will help ensure best practices and maximise the accuracy of results.

Conflict of interest statement

All claims made in this article are solely those of the authors and do not necessarily reflect the views of their affiliated company or organisation.

References

[1] R. Szibor, M. Krawczak, S. Hering, et al., Use of X-linked markers for forensic purposes, Int. J. Legal Med. 117 (2003) 67–74.

[2] R. Szibor, X-chromosomal markers: Past, present and future, Forensic Sci. Int. Genet. 1 (2007) 93–99.

[3] A.O. Tillmar, D. Kling, J.M. Butler, et al., DNA Commission of the International Society for Forensic Genetics (ISFG): Guidelines on the use of X-STRs in kinship analysis, Forensic Sci. Int. Genet. 29 (2017) 269-275.

[4] L. Cainé, R. Carvalho, S. Costa, et al., Interest of X chromosome (Argus X-12 kit) in complex kinship analysis, Forensic Sci. Int. Genet. Suppl. Ser. 3 (2011) e206-e207.

[5] M. Scherer, M. König, M. Bussmann, et al., Development and validation of the new Investigator® Argus X-12 QS Kit, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e256-e257.

[6] S. Iyavoo, S. Knights, M. Mavrommatis, et al., Implementation of Prep-n-Go™ Buffer for DNA extraction from buccal swabs. Forensic Sci. Int. Genet. Suppl. Ser. 7 (2019) 211-213.

[7] J. Perry, T. Munshi, T. Haizel, et al., Validation of reduced volume VeriFiler™ Express PCR Amplification Kit for buccal swab samples extracted using Prep-n-Go™ Buffer. J. Forensic Sci. 67 (2022) 1971-1978.

[8] C. Gomes, M. Magalhães, A. Amorim, et al., How useful is your X in discerning pedigrees? Forensic Sci. Int. Genet. Suppl. Ser. 3 (2011) e161-e162.

[9] S. Hering, J. Edelmann, S. Haas, et al., Paternity testing of two female siblings with Investigator Argus X-12 kit: A case with several rare mutation and recombination events, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e341–e343.

[10] M. Nothnagel, R. Szibor, O. Vollrath, et al., Collaborative genetic mapping of 12 forensic short tandem repeat (STR) loci on the human X chromosome, Forensic Sci. Int. Genet. 6 (2012) 778-784.

[11] I. Gomes, N. Pinto, S. Antão-Sousa, et al., Twenty Years Later: A Comprehensive Review of the X Chromosome Use in Forensic Genetics, Front. Genet. 11 (2020) 926.

[12] A.R. Alwi, N.A. Mahat, F.M. Salleh, et al., Population data for 12 X-STRs loci in Malaysian Malay and Chinese populations, Heliyon (2024) e38054.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 385-391
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Simplified Differential Extraction (DE) Method for the RapidHIT™ ID to Obtain Investigative Leads from Sexual Assault Evidence

Daniella Gonzalez Tirado[1], Brooke Bass[1], Jason Werking[2], Rob Lagacé[2], Jack Ballantyne[1,3*], Erin Hanson[1,3,4*]

[1] *National Center for Forensic Science, University of Central Florida, Orlando, FL, USA.*
[2] *ThermoFisher Scientific, Pleasanton, CA, USA.*
[3] *Department of Chemistry, University of Central Florida, Orlando, FL, USA.*
[4] *Violence Against Women faculty cluster, University of Central Florida, Orlando, FL, USA.*

** Corresponding author at: University of Central Florida, 12354 Research Parkway Ste 225, Orlando, FL, USA, erin.hanson@ucf.edu, +1 (407) 823 0642.*

Abstract

Rapid DNA integrates sample analysis workflow from extraction to capillary electrophoresis, generating autosomal STR profiles in as little as 90 minutes without the need for a DNA laboratory. With the ability to utlize this technology forthcoming (e.g. by early 2025), the purpose of the current work was to evaluate simplified off-instruments methods for processing sexual assault (SA) evidence for subsequenct use with the RapidHIT™ ID system allowing investigative leads for law enforcement to be generated in shorter amount of time tan standard SA evidence processing workflows. Here we report the use of a simplified differential extraction (DE) method that provides a sperm-enriched fraction that can be run on the RapidHIT™ system. We demonstrate the successful recovery of single source male (i.e. sperm donor) profiles from mock semen-epithelial (buccal and vaginal, to simulate both oral and sexual assaults) mixtures with as little as 1 ml of semen and from *bona fide* volunteer provided cervicovaginal post-coital and drainage samples collected 12 – 60 hours after intercourse.

Keywords

Rapid DNA, RapidHIT™ ID, sexual assault evidence, differential extraction.

1. Introduction

Rapid DNA integrates sample analysis workflow from extraction to capillary electrophoresis, generating autosomal STR profiles in as little as 90 minutes without

the need for a DNA laboratory. Currently, the FBI allows profiles developed from *reference samples* on rapid DNA instruments to be uploaded to CODIS, but not from *crime scene samples* [1]. Reference samples collected from qualifying arrestees during booking are enrolled and automatically searched against unsolved crimes. This is expanding nationwide due to its ability to link arrestees to unrelated crimes while still in custody.

Although the FBI does not currently allow DNA profiles generated from crime scene samples on rapid DNA instruments to be uploaded to CODIS, they are working with vendors to enable this capability in early 2025. In the meantime, many law enforcement agencies have been using rapid DNA outside of CODIS to substantially impact criminal investigations, human trafficking, and the identification of human remains [2,3]. While most of this work has been done with blood and saliva cases, rapid DNA is rarely used for sexual assault cases because rapid DNA instruments do not perform differential extraction [4,5]. We sought to develop an off-instrument differential extraction method that was compatible with non-technical users and rapid DNA instruments to enable law enforcement to take advantage of the speed of rapid DNA in sexual assault cases.

The outcome is a simple workflow that utilizes a 1-hour differential lysis to preferentially lyse epithelial cells leaving sperm cells intact. After a few brief washes, the epithelial cell fraction is separated, and the remaining sperm pellet is briefly incubated with 1-thioglycerol. Five microliters of the 30 ml sperm fraction is added to a HydraFlock® swab and run on the RapidHITTM ID system (low lysis volume protocol). Single source male DNA profiles were obtained from as little as 1 ml of semen, with an in-depth evaluation of the method's sensitivity still being evaluated. Mock mixtures using both buccal and vaginal epithelial cells were tested representing possible casework scenarios of vaginal and oral assaults. Successful results were also obtained from volunteer donated 12-60 hr post coital simples in addition to a post-coital drainge(i.e. underwear) sample.

2. Material studied, methods, techniques

2.1. Body fluid sample preparation

All body fluid samples were collected in accordance with the procedures approved by the University's Institutional Review Board. Buccal swabs were collected from male and female volunteers by swabbing the inside of the cheek. Single source semen samples were purchased from a commercial source (BioIVT, Westbury, NY). Mock mixture samples were prepared using a buccal or vaginal secretions swab

and pipetting liquid semen onto the swab in specified volumes. Post-coital cervico-vaginal swabs (x2 for each time point) were collected by a female volunteer who recovered the samples after separate acts of sexual intercourse at various time points (12, 24, 48 and 60 hours). The volunteers were instructed to take the samples from the cervix by swabbing multiple times for 20-30 sec at each specific time interval. All volunteers were asked to abstain from intercourse at least 2 to 3 days prior to sample collection. A post-coital drainage sample was collected 13 hours after vaginal intercourse by collecting the volunteer's underwear. All samples were dried overnight and stored at -20°C until analysis.

2,2. Simplified differential extraction (DE)

A whole cotton swab (i.e. swab head) was placed in a Prepfiler™ LySep™ sample tube to which 400 ml of stain extraction buffer (50mM NaCl, 10mM Tris-HCl, 10mM EDTA, 2% SDS) and 10 ml proteinase K (10 mg/mL) was added. A Prepfiler™ LySep™ column was placed in the sample tube and the sample was incubated at 56°C for 1 hour (900 rpm). The swab head was then placed inside the LySep™ column, placed into the original sample tube and centrifuged at 13,000 rpm for 5 min. The supernatant (nsp; non-sperm fraction) was transferred to a new sterile 1.5 mL tube and stored at -20°C. The remaining sperm pellet was washed (2x) with 250 ml stain extraction buffer. The washed sperm pellet was resolubilized in 30 ml stain extraction buffer. One microliter of 1-thioglycerol was added and the sample was incubated at room temperature for 5 minutes. Samples were used immediately for analysis or stored at 4°C until needed.

3.1. Rapid DNA analysis

Five microliters of the prepare sperm fraction (see section 2.2) were pipetted onto a HydraFlock® sterile swab (Puritan®). The swab was placed into a RapidINTEL™ Plus sample cartridge. Swabs were prepared immediately prior to analysis, however in some instances the swab head was placed inside a 1.5 mL tube, dried at room temperature overnight and stored at 4°C until needed. All samples were run on the RapidHIT™ system (AB RapidHIT™ ID v2.0) using the low lysis volume protocol. All profiles were analyzed using RapidLink™ v2.0 software.

3. Results

Mock mixtures utilizing buccal and vaginal epithelial cells mixed with semen (1 ml) were prepared to simulate both oral and sexual assaults. The admixed samples

were analyzed using the developed simplified DE method. Full single source male profiles were obtained for all samples with no female allele carry over or male allele-drop out.

Successful results were also obtained from volunteer donated 12 – 60 hr cervicovaginal samples as well as from a post-coital drainage underwear sample collected 13 hrs after intercourse. For the 18 hr post-coital sample a full male profile was obtained with only four alleles attributable to the female donor. For the 36 hr post-coital sample, a full male profile was also obtained with no alleles attributable to the female donor present. The second swab from each of these post-coital samples was extracted using a non-differential PrepFiler Express™ extraction (30 ml elution) and quantitated with the Quantifiler™ Trio kit in order to obtain an approximate F/M (female/male) ratio for the samples. A F/M ratio of ~16:1 and ~37:1 was obtained for the 18 and 36 hr samples, respectively. Using the simplified DE rapid DNA approach developed here, a full single source male profile was obtained even with a ~37:1 F/M ratio.

For the 36 hr post coital sample, a full male profile was obtained with only two alleles attributable to the female donor observed (Fig 1A). A pair of underwear worn by the female participant was collected 13 hours after intercourse and a small cutting from the underwear was evaluated using the developed simplified DE method. As can be seen in Figure 1B, a full single source male DNA profile was obtained with no alleles attributable to the female observed. All other post-coital samples tested were cervicovaginal samples collected using sterile cotton swabs. The post-coital drainage sample demonstrates the ability to obtain results from cloth as well as swab samples.

A 60 hr post coital sample was also tested. A probative male profile was obtained from the 60 hr sample with drop-out of only two male alleles observed. However, an increased number of alleles (8 in this instance) attributable to the female donor were observed in this sample. We are currently working to develop protocol modifications that may further improve removal of residual non-sperm alleles prior to rapid DNA analysis. With 25 of the 30 ml sperm fraction remaining there is sufficient sample for additional steps or further sample purification.

4. Discussion

While rapid DNA is currently being utilized by law enforcement to develop investigative leads in criminal investigations, human trafficking and identification of

human remains, in a forensic lab personnel can also utilize this technology to process urgent samples more efficiently. One example of this would be the analysis of sexual assault evidence in order to identify a potential perpetrator in a more timely manner. With the ability to upload crime scene samples to CODIS expected in 2025, it will be important to have protocols in place to ensure the successful processing of this type of evidence. The results of this study demonstrate the ability to relatively seamlessly integrate rapid DNA into existing workflows while leveraging the speed of analysis to permit rapid profile recovery.

In the developed simplified DE method, not only is sufficient male DNA obtained to generate full single source autosomal STR profiles, sufficient sample is retained in several stages of the process to permit additional sample re-analysis using both rapid DNA and standard laboratory workflows. For standard two-swab evidence collection, one full swab would be retained for future analysis. The second swab would be used in the developed simplified DE protocol. From this process, a full non-sperm fraction as well as the original swab substrate, which often contains residual sperm that were not eluted from the cotton during initial processing, are collected and stored. Only 5 of 30 ml of sperm fraction is used in the initial sample run, leaving over 80% of the sperm fraction available for future analysis. Additionally, in many instances the original swab analyzed by rapid DNA analysis can be re-tested in a second rapid DNA cartridge or re-tested using standard DNA extraction workflows with probative information still obtained.

Currently the developed protocol is intended for use by laboratory personnel. The goal of future work would be to continue to optimize the protocol to permit transition from a lab setting to a point-of-collection site to further increase timeliness and generate investigative leads potentially while a sexual assault exam is still being performed.

5. Conclusion

Initial studies have successfully demonstrated the ability to utilize a simplified off-instrument differential extraction (DE) method to process sexual assault evidence prior to use of rapid DNA analysis. The results shown here demonstrate the potential of this approach to aid law enforcement in faster development of investigative leads and possible identification of perpetrators. Future work will be focused on obtaining additional volunteer donor couples to further evaluate the developed protocols. A possible collaborative study with other agencies and universities will

even further expand the validation testing of this approach and to facilitate implementation of these protocols into other laboratories.

6. Acknowledgments

The authors would like to acknowledge all anonymous donors who provided samples for this study. This work was supported through a Sponsored Research Agreement (SRA) from ThermoFisher Scientific. The sponsoring agency (as co-authors on the study) had a role in analysis and interpretation of data and in the editing of the manuscript.

7. Conflict of interest statement

No conflicts of interest.

8. References

[1] Scientific Working Group on DNA Analysis Methods (SWGDAM, Scientific Working Group on DNA Analysis Methods Position Statement on Rapid DNA Analysis,https://www.swgdam.org/_files/ugd/4344b0_f84df0465a2243218757fac1a1ccffea.pdf. (2017) 1-3.

[2] M. Eaton, M. S. Woolf, S. Pemmasani, et al., Two simplified toot sample preparation methods for conventional laboratory and RAPIDHITTM ID workflows, J. Forensic Sci. (2024) doi.org/10.1111/1556-4029.15624.

[3] J. Watherston, J. Watson, D. Bruce, et al., An in-field evaluation of rapid DNA instruments for disaster victim identification, Int J Legal Med. 136 (2) (2022) 493-499.

[4]. P. Gill, A.J. Jeffreys and D.J. Werrett, Forensic application of DNA 'fingerprints,' Nature. 318 (1985) 577-579.

[5]. C. Clark, R. Turiello, R. Cotton, et al., Analytical approaches to differential extraction for sexual assault evidence, Analytica Chimica Acta. 1141 (2021) 230-245.

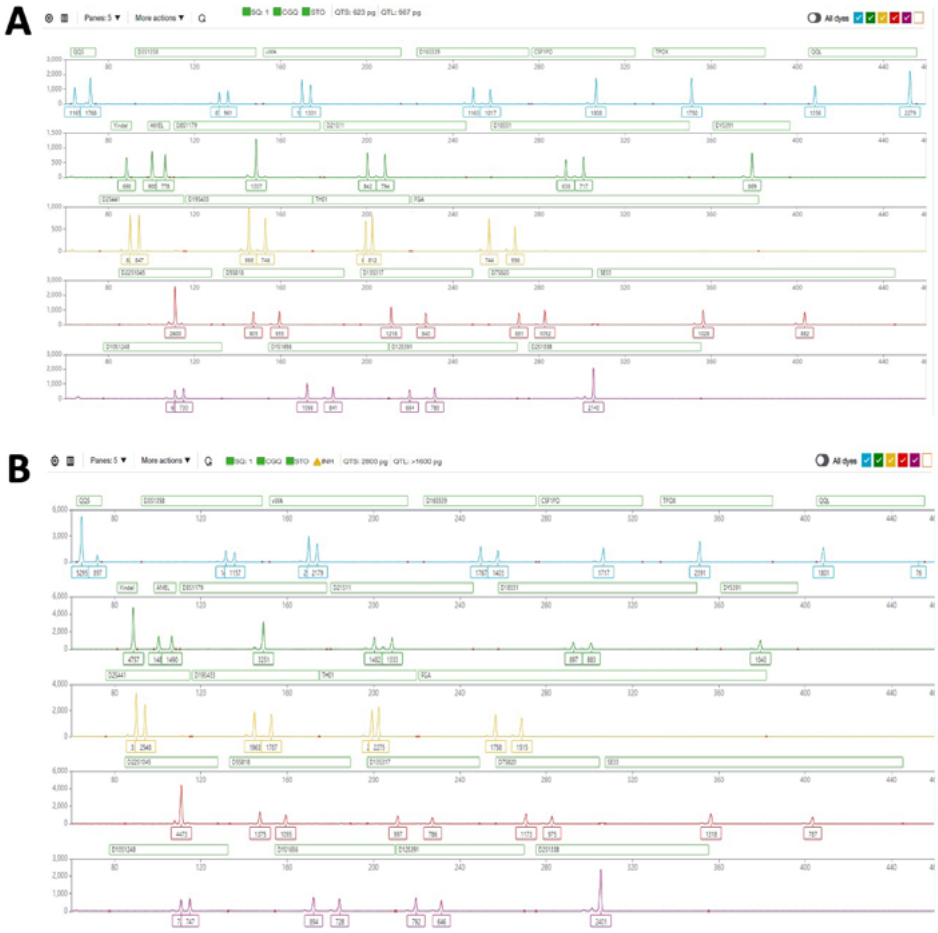


Figure 1. Single source male autosomal STR profiles obtained from post-coital samples using a simplified DE method for analysis of sexual assault evidence on the RapidHIT™ ID system. A) A cervicovaginal sample collected 36 hours after intercourse was analyzed with the developed simplified differential extraction method and was run on the RapidHIT™ ID system using a RapidINTEL™ Plus cartridge. A full male profile was obtained with no female allele carry-over. B) A post-coital drainage sample was obtained from volunteer collected underwear 13 hours after intercourse. Using the same simplified DE method, a full male profile was also obtained with no female allele carry-over. For both samples, profile accuracy was determined by comparison to reference profiles. X-axis: size, in base pairs (bp). Y-axis: relative fluorescence units (RFUs).

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 393-401
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Exploring Immunomagnetic Isolation of Spermatozoa

Julia Wang, MS[1*], Samantha Davis, MS[1], Sheree Hughes, PhD[1], Brendan Chapman, PhD[2], Andrew Currie, PhD[2], Rachel Houston, PhD[1]

[1] *Department of Forensic Science, Sam Houston State University, Huntsville, USA*

[2] *Medical, Molecular and Forensic Science, Murdoch University, Perth, Australia*

**Corresponding author at: 1003 Bowers Blvd. Huntsville, TX 77340, Julia.Wang@shsu.edu, Fax +1-936-294-4905*

Abstract

Efficient, high-throughput methods are essential for processing sexual assault kits (SAKs), which typically contain a mixture of female epithelial cells and male spermatozoa. This study tested sperm-specific antibodies conjugated to magnetic beads for isolating sperm. Immunocytochemistry confirmed the expression of target proteins and antibody binding; however, the magnetic beads also showed binding affinity for both sperm and vaginal epithelial (VE) cells, even without antibodies and despite the use of blocking agents. These non-specific interactions increased with longer incubation times. While previous studies have reported mixed success with immunomagnetic isolation, this work highlights the need to address non-specific binding and cross-reactivity to fully realize the potential of antibody-conjugated beads in differential cell isolation. While immunological techniques are employed in forensic laboratories for screening, caution is warranted when considering their broader application for cell isolation in SAK processing.

Keywords

Sexual assault, Spermatozoa, Differential Extraction, Antibodies, Forensic DNA

Introduction

Forensic DNA laboratories often process sexual assault kits, containing intimate swabs that predominantly contain female epithelial cells and few sperm cells from male perpetrators. Separating these cell types for DNA extraction is labor-intensive and time-consuming, and not always effective.

Current methods for testing SAK evidence, including differential lysis extraction, separate sperm from epithelial cells but can result in a significant loss of male DNA during processing. Alternative techniques, such as laser microdissection, acoustic cell trapping, and flow cytometry, face challenges related to cost and implementation in forensic laboratories. This study investigates the effectiveness of magnetic beads conjugated to sperm-specific antibodies (SPAM1, SPACA1, ZPBP1) for isolating sperm cells from mixtures [1–3]the most common evidence gathered comes in the form of intimate swabs. These swabs contain cells from both the female and male. In these cases the most common cell present is epithelial cells from the female with relatively few spermatozoa from the male. To produce an STR DNA profile suitable for an investigation, separation of the cell types is required to prevent a mixed profile that may complicate interpretation. While there are methods that can achieve this separation, they are labour intensive and have no guarantee of complete separation of the spermatozoa. Immunomagnetic bead separation techniques are being adapted from other scientific fields to be used in forensic biology to separate spermatozoa from mixed substrates. This study tested the binding of a Hyaluronidase PH-20 (SPAM1. By optimizing conditions for sperm isolation, the research aims to enhance the processing of SAK evidence and improve the detection of sperm DNA in complex samples.

Material studied, methods, techniques

Samples

Samples were collected from informed and consenting volunteers using the approved Sam Houston State University Institutional Review Board protocol IRB-2020-248. Reference buccal swabs were collected from all volunteers. Vaginal swabs from four female donors and seminal fluid from four male donors were collected.

VE cells and sperm cells were counted using a hemacytometer [4]. Cell suspensions were prepared at various concentrations (103, 104, 105, 106 cells per assay) to test assay specificity and sensitivity.

Immunocytochemistry

Sperm cell suspensions (10^6 cells/mL) were spotted onto glass slides, dried at 60°C for 30 seconds, and then fixed using 4% formaldehyde for 15 minutes at room temperature. After washing three times with PBS, the fixed cells were blocked with 5% normal goat serum (Jackson ImmunoResearch Laboratories Inc., West Grove,

PA, USA) for 2 hours at room temperature. Sperm cells were then incubated overnight at 4°C in a humidity chamber with 1:200 anti-SPAM1 antibody, 1:500 anti-SPACA1 antibody, 1:200 anti-ZPBP1 antibody, or a cocktail of all three antibodies (Thermo Fisher Scientific, Waltham, MA, USA). After three PBS washes, the slides were incubated with Alexa Fluor 488 conjugated goat anti-rabbit secondary antibody (Invitrogen, Carlsbad, CA, USA) diluted 1:500 for 1 hour at room temperature. Nuclei were stained with DAPI (Thermo Fisher Scientific) for 10 minutes at room temperature. Slides were mounted with Entellan™ (Sigma-Aldrich, St. Louis, MO, USA), and negative controls were processed without primary antibodies. Slides were examined with a Leica FS CB compound microscope at 100x magnification (Leica Microsystems, Wetzlar, Germany).

Ligand-Bead Coupling

Dynabeads™ M-450 Epoxy

Dynabeads® M-450 Epoxy (Thermo Fisher Scientific) were prepared according to the manufacturer's guidelines and blocked with 0.1% BSA. Antibody-bead coupling was performed using Buffer 1 (0.1 M sodium phosphate buffer, pH 7.4), while Buffer 2 (Ca^{2+} and Mg^{2+} free PBS supplemented with 0.1% BSA and 2 mM EDTA) was used for washing cells and cell isolation. The antibody cocktail was coupled onto the beads at 0.025 mg/mL anti-SPAM1, 0.020 mg/mL anti-SPACA1, and 0.002 mg/mL anti-ZPBP. No antibodies were added to the control beads to assess non-specific isolation. The antibodies and beads were incubated together for 24 hours at room temperature while rotating at 15 rpm using the HulaMixer Sample Mixer (Thermo Fisher Scientific).

Dynabeads™ M-270 Carboxylic Acid

Dynabeads™ M-270 Carboxylic Acid (Thermo Fisher Scientific) were prepared following the manufacturer's two-step procedure with N-hydroxyl succinimide. To characterize the potential cell-binding properties of the beads without antibody conjugation, M-270 beads were either activated and quenched or activated and coated with 0.1% bovine serum albumin (BSA) (Thermo Fisher Scientific) for blocking.

Antibody Quantification

The antibody concentration in the supernatant washes from the preparation of antibody-coated M-450 Epoxy beads were quantified by Pierce™ BCA Protein Assay kit (Thermo Fisher Scientific) to assess the efficiency of the bead-antibody

conjugation. Reactions were measured using a BioTek Epoch microplate spectrophotometer (Agilent Technologies, Inc., Santa Clara, CA, USA).

Cell Isolation

Cell Isolation with Dynabeads™ M-450 Epoxy

For cell isolation with the M-450 Epoxy beads, 10^5 of sperm cells or VE cells were incubated with either antibody-conjugated or BSA control beads for 15, 30, 60, or 120 minutes with 15 rpm rotation at 4°C. The samples were applied to a magnet for 2 minutes. The supernatant fraction was collected for extraction, and the bead-cell complexes were washed four times with Buffer 2 to remove unbound cells prior to extraction. To determine the sensitivity of the antibody cocktail beads, sperm cells were applied to either antibody-conjugated or BSA control beads for 30 minutes at different amounts (10^3, 10^4, 10^5, 10^6). As with the incubation time study, both bead and supernatant fractions were collected for extraction.

Cell Isolation with Dynabeads™ M-270 Carboxylic Acid

To isolate sperm cells and VE cells from single source samples, cell suspensions were counted and applied to the prepared M-270 Carboxylic Acid beads for 1 hour at room temperature with 15 rpm rotation. The samples were applied to a magnet (Invitrogen) for 2 minutes. The supernatant fraction was collected for extraction, and the bead-cell complexes were washed three times with PBS.

DNA Extraction and Quantification

Supernatant and bead fractions from cell isolation were lysed, purified, and eluted in 50 µL using the QIAamp® DNA Investigator kit (QIAGEN, Hilden, Germany) following the manufacturer's protocol for total DNA isolation from sexual assault specimens. Following lysis, bead fractions were placed on a magnet for 2 minutes, and the lysis supernatants were collected and purified.

DNA extracts were quantified using Quantifiler Trio DNA Quantification kit (Thermo Fisher Scientific) on an Applied Biosystems 7500 real-time PCR instrument (Thermo Fisher Scientific) following the manufacturer's recommendations.

Results and Discussion

Visualization of antibody detection for the target proteins indicated that signals for SPAM1 and SPACA1 were more intense than those of ZPBP (Figure 1). Additionally, the antibody cocktail effectively detected protein expression.

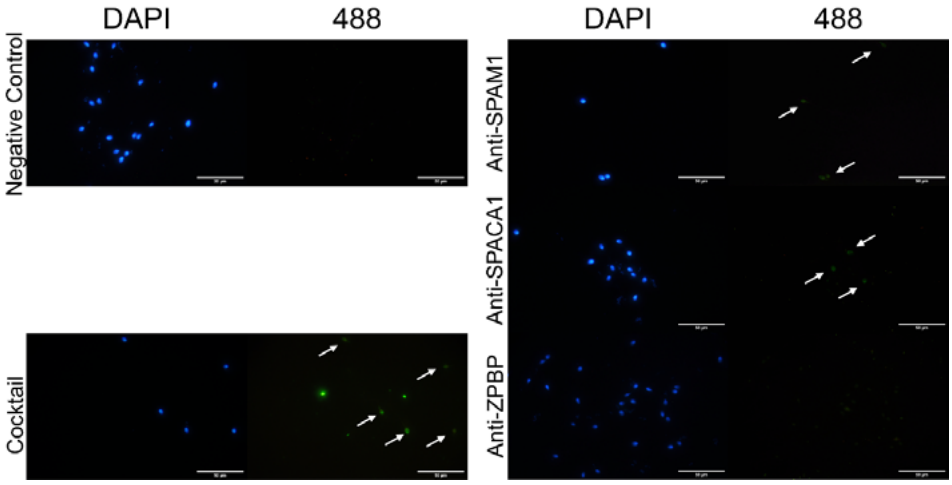


Figure 1. Immunocytochemistry characterization of spermatozoa SPAM1, SPACA1, and ZPBP expression. Arrows indicate spermatozoa with detected signal from secondary antibody in SPAM1, SPACA1, and antibody cocktail samples. All slides were stained with DAPI to identify cell nuclei, shown in blue. All primary antibodies were detected with the secondary antibody conjugated with Alexa Fluor 488, shown in green. The negative control slide was prepared without primary antibody.

A cocktail of anti-SPAM1, anti-SPACA1, and anti-ZPBP antibodies was conjugated to M-450 beads to evaluate potential additive effects in sperm cell isolation. Antibody quantification by BCA assay determined that in the cocktail antibodies, an average of 35% of the antibodies added during the bead preparation were retained on the bead surfaces. When incubated with both sperm and VE cells, DNA extracted from the bead fractions increased with incubation time (Figure 2A). This trend in both antibody cocktail and BSA beads suggests the increase in extracted DNA could be attributed to higher amounts of non-specific binding.

Sensitivity tests showed that *the incorporation of a sperm-targeted antibody cocktail did not enhance spermatozoa binding compared to beads treated with only BSA (Figure 2B). Although antibody binding to the M-450 Epoxy beads was confirmed, potential issues with the conformation of the coupled antibodies may have hindered effective sperm isolation. Addressing these challenges will be crucial for optimizing immunomagnetic isolation techniques in future research.*

This study identified non-specific binding of both spermatozoa and VE cells to magnetic beads as a significant challenge in immunomagnetic isolation-based

differential extraction. Using Dynabeads™ M-450 Epoxy, designed for cell isolation, demonstrated non-specific binding. As spermatozoa are smaller than the epithelial or blood cells typically targeted by bead-based methods, using smaller magnetic beads, such as nanoparticles with higher binding capacities, may enhance aggregation to sperm cells and improve isolation outcomes [5].

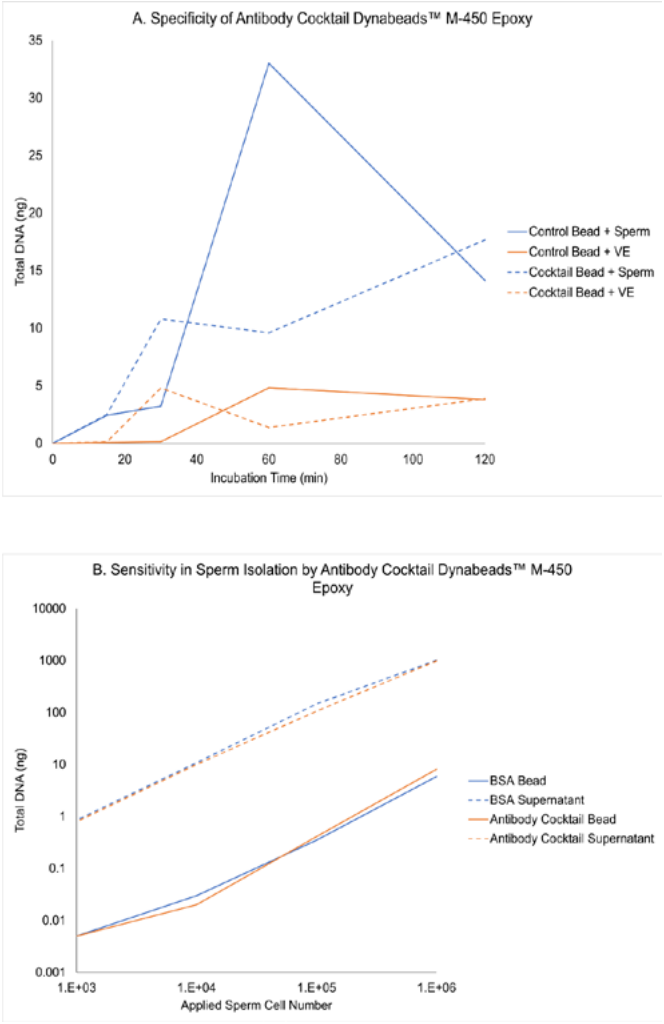


Figure 2. A) Total DNA extracted from 0.1% BSA or antibody cocktail-conjugated M-450 magnetic beads increased with incubation time for both spermatozoa and VE cell types. B) Total DNA extracted from 0.1% BSA or antibody cocktail bead and supernatant fractions.

As an alternative to the M-450 Epoxy beads, the smaller M-270 Carboxylic Acid beads, shown previously to isolate sperm cells from mixtures, were tested for non-specific binding to spermatozoa and VE [6]which is a glycosylphosphatidylinositol-anchored hyaluronidase located on the head of sperm, has important functions in fertilization. Here we describe a newly developed method for sperm isolation using anti-PH-20 antibody-coupled immunomagnetic beads (anti-PH-20 IMBs. Without antibodies, DNA recovery was 12.5% for spermatozoa and 18.4% for VE cells. With 0.1% BSA as a blocking agent, recovery from spermatozoa dropped to 2.1%, but non-specific binding to VE cells increased, with DNA recovery rising to 25.5%. Although a previous study has reported *success in isolating spermatozoa from cell mixtures* [6]which is a glycosylphosphatidylinositol-anchored hyaluronidase located on the head of sperm, has important functions in fertilization. Here we describe a newly developed method for sperm isolation using anti-PH-20 antibody-coupled immunomagnetic beads (anti-PH-20 IMBs*, the use of a blocking agent did not mitigate the non-specific binding of VE cells to the Dynabeads™ M-270 Carboxylic Acid in our study.*

While several studies have demonstrated varying degrees of success with immunomagnetic spermatozoa isolation, the intrinsic properties of magnetic beads that contribute to non-specific binding have been largely overlooked [6–9] which is a glycosylphosphatidylinositol-anchored hyaluronidase located on the head of sperm, has important functions in fertilization. Here we describe a newly developed method for sperm isolation using anti-PH-20 antibody-coupled immunomagnetic beads (anti-PH-20 IMBs. *However, one sperm isolation study by Becks did report similar findings of non-specific interactions* [10]. *Moreover, BSA, commonly used as a blocking agent, has been shown to exhibit cross-reactivity with target proteins, including antibodies* [11]we found nonspecific binding of VCP to BSA and identify a BSA preparation that did not result in non-specific binding. This work draws attention to the fact that not all BSA preparations are alike. It also highlights the need to perform critical controls to ensure that ELISA reactants do not inappropriately bind to the blocking agent.","container-title":"Journal of immunological methods","DOI":"10.1016/j.jim.2012.06.009","ISSN":"0022-1759","issue":"1-2","journalAbbreviation":"J Immunol Methods","note":"PMID: 22732194\nPMCID: PMC3432671","page":"148-151","source":"PubMed Central","title":"Enzyme-Linked Immunosorbent Assay (ELISA. *To address this, alternative blocking agents such as serum, gelatin, or non-protein reagents could be explored to minimize non-specific interactions and clarify the extent of BSA's cross-reactivity.*

Conclusion

This study sought to advance magnetic bead-based spermatozoa isolation but underscores the need for a more thorough understanding of the interactions between beads, antibodies, and target cells before these techniques can be applied in forensic settings. To make immunomagnetic isolation viable for analyzing SAK evidence, significant progress is needed in enhancing the specificity of the bead-antibody complexes and optimizing isolation protocols. Continued research is essential to overcome these limitations and fully realize the potential of immunomagnetic sperm isolation in forensic science.

Acknowledgments

This work was financially supported by the Forensic Sciences Foundation Lucas Research Grant.

Conflict of interest statement

None.

References

[1] S. Hutchinson, B. Chapman, G. Turbett, Sperm fishing: antigen-antibody complexes for the capture and enrichment of spermatozoa in mixed cell substrates, Australian Journal of Forensic Sciences 51 (2019) S95–S98. https://doi.org/10.1080/00450618.2019.1573926.

[2] A.N. Yatsenko, D.S. O'Neil, A. Roy, P.A. Arias-Mendoza, R. Chen, L.J. Murthy, D.J. Lamb, M.M. Matzuk, Association of mutations in the zona pellucida binding protein 1 (ZPBP1) gene with abnormal sperm head morphology in infertile men, Mol Hum Reprod 18 (2012) 14–21. https://doi.org/10.1093/molehr/gar057.

[3] T. Zhang, J. Wang, W. Niu, F. Wang, J. Liu, Y. Xing, P. Jia, X. Ren, J. Wang, W. Zang, X. Chen, Bioinformatic prediction of the structure and characteristics of human sperm acrosome membrane-associated protein 1 (hSAMP32) and

evaluation of its antifertility function in vivo, Reprod Fertil Dev 32 (2020) 1282–1292. https://doi.org/10.1071/RD20198.

[4] M.R. Green, J. Sambrook, Estimation of Cell Number by Hemocytometry Counting, Cold Spring Harbor Protocols 2019 (2019) pdb.prot097980. https://doi.org/10.1101/pdb.prot097980.

[5] L. Borlido, A.M. Azevedo, A.C.A. Roque, M.R. Aires-Barros, Magnetic separations in biotechnology, Biotechnology Advances 31 (2013) 1374–1385. https://doi.org/10.1016/j.biotechadv.2013.05.009.

[6] X.-C. Zhao, L. Wang, J. Sun, B.-W. Jiang, E.-L. Zhang, J. Ye, Isolating Sperm from Cell Mixtures Using Magnetic Beads Coupled with an Anti-PH-20 Antibody for Forensic DNA Analysis, PLOS ONE 11 (2016) e0159401. https://doi.org/10.1371/journal.pone.0159401.

[7] K. Anslinger, B. Bayer, S.M. Danilov, R. Metzger, Application of sperm-specific antibodies for the separation of sperm from cell mixtures, Forensic Science International: Genetics Supplement Series 1 (2008) 394–395. https://doi.org/10.1016/j.fsigss.2007.10.140.

[8] X.-B. Li, Q.-S. Wang, Y. Feng, S.-H. Ning, Y.-Y. Miao, Y.-Q. Wang, H.-W. Li, Magnetic bead-based separation of sperm from buccal epithelial cells using a monoclonal antibody against MOSPD3, Int J Legal Med 128 (2014) 905–911. https://doi.org/10.1007/s00414-014-0983-3.

[9] X. Li, F. Xu, J. Zheng, M. Sun, X. Zhu, P. Lv, Z. Du, X. Zhang, J. Yao, Magnetic bead-based separation of sperm cells from semen-vaginal fluid mixed stains using an anti-ACRBP antibody, Int J Legal Med 137 (2023) 511–518. https://doi.org/10.1007/s00414-022-02917-8.

[10] K. Becks, Immunomagnetic beads coupled with anti-PH20 antibodies to isolate sperm from DNA mixtures for sexual assault kit application, Thesis, West Virginia University, 2019.

[11] Y. Xiao, S.N. Isaacs, Enzyme-Linked Immunosorbent Assay (ELISA) and Blocking with Bovine Serum Albumin (BSA) - Not all BSAs are alike, J Immunol Methods 384 (2012) 148–151. https://doi.org/10.1016/j.jim.2012.06.009.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 403-409
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Single cell sorting of spermatozoa: a comparative study with differential lysis for integration in forensic casework

Kristina Fokias[1,2], Iene Rutten[2], Jeroen Lammertyn[2], Bram Bekaert[1,3*]

[1]*KU Leuven, Department of Imaging and Pathology, Forensic Biomedical Sciences, Leuven, Belgium*
[2]*KU Leuven, Department of Biosystems, Biosensors group, Leuven, Belgium*
[3]*University Hospital UZ Leuven, Laboratory of Forensic Genetics, Leuven, Belgium*

** Corresponding author at: KU Leuven, Forensic Biomedical Sciences, Herestraat 49 – box 7003 71, 3000 Leuven, Belgium · Bram.bekaert@uzleuven.be*

Abstract

A vital aspect in the identification of sexual assault offenders often entails the detection and separation of spermatozoa in samples collected from the victim. Differential lysis is the conventional method for the isolation of spermatozoa, though is time-consuming and often inadequate to produce male DNA profiles in samples containing an unfavourable sperm/epithelial cell ratio. The aim of this study was to develop a methodology for analysing samples with a low number of spermatozoa, such as in cases where there is a long interval between the alleged sexual assault and sample collection. Fluorescence-activated cell sorting (FACS) is a well-established technology often used in the clinical setting to enrich rare target cells from a mixed cell populations based on phenotypical characteristics detected by flow cytometry. Although its application in forensics has previously been reported on a limited basis, reliable validation is fundamental for its implementation and use in forensic casework. With this research, a methodology has therefore been established to sort sperm from epithelial cells by means of FACS, which is followed by a direct lysis and PCR protocol for subsequent DNA profiling. The performance of this method was then assessed and compared with standard differential lysis through parallel testing of samples with different sperm/epithelial cell ratios and post-coital swabs. These analyses indicate an improved separation efficiency and a higher DNA profile quality for the new method, making it a viable alternative to differential lysis. Upon completion of the validation, this methodology will be implemented for the analysis of casework samples where differential lysis is expected to fail. As such, the implementation of more sensitive methods could improve the success rate of identifying perpetrators and therefore potentially have a major impact on the conviction rate of sexual offenders.

Keywords
Flow cytometry, FACS, forensic DNA, sexual assault, differential lysis, sperm.

1. Introduction

Differential lysis (DL) is the conventional method for the isolation of spermatozoa in sexual assault casework. However, several drawbacks are associated with this technique, including the inability to produce male DNA profiles in samples containing an unfavourable sperm/epithelial cell (SC/EC) ratio. With as much as 90% of male DNA lost in a typical DL process [1], alternative approaches for the enrichment of sperm cells have been widely investigated [2]. One such alternative is by means of fluorescence-activated cell sorting (FACS), which relies on the flow cytometric detection of cellular phenotypical characteristics. Despite its extensive application in the clinical setting, only a handful of research articles are available on its use in forensics (*e.g.,* [3,4]). In particular, a thorough comparison of DL and FACS employing the most recent technologies is missing. However, such benchmarking studies are fundamental to explore the full potential of new methods. Following the promising results of our preliminary study in which a sperm sorting procedure was established [5], its performance was compared with DL through parallel testing of assembled mixtures and post coital (PC) samples. A positive evaluation of this comparison will form the basis for a systematic validation considering the sensitivity, specificity, reproducibility, and repeatability of the assay.

2. Material studied, methods, techniques

To compare the sensitivity of the FACS sorting method with DL, cell suspensions constituted of 500,000 vaginal EC and a declining number of SC were analysed in triplicate using both methods. In addition, six couples volunteered to donate PC swabs collected in duplicate up to 72 hours following intercourse. For sperm sorting, the samples were stained using the Sperm Hy-Liter™ SOS kit and sorted using a Sony MA900 Multi-Application cell sorter (Sony Biotechnology Inc.). Positive events were sorted in a 96 well plate containing 5 µl of lysis buffer (0.1 mg/ml proteinase K, 5 mM DTT) per well and subjected to DNA amplification (30 cycles) using the PowerPlex® Fusion system (Promega). DNA amplicons were separated using an ABI 3730xl instrument (Applied Biosystems) and the resulting STR profiles analysed with GeneMarker® HID software (SoftGenetics, 3.0.0).

Promega's Tissue and Hair Extraction kit was used for DL. Briefly, the samples were centrifuged (12,000 rpm, 10 minutes), the supernatant removed and the pellet resuspended in 100 µl incubation buffer/proteinase K solution. Following a 2-hour incubation step at 56 °C, 300 µl lysis buffer was added prior to a second centrifugation step (13,000 rpm, 5 minutes). The resulting supernatant (= female fraction) was transferred to another tube. One hundred µl incubation buffer/proteinase K/DTT solution was then added to the remaining cell pellet (= male fraction) for the second incubation step (56 °C, 1 hour), followed by the addition of 300 µl lysis buffer/DTT solution. The DNA IQ™ Casework Pro kit and PowerQuant® System were used for subsequent DNA extraction and quantification. Five µl purified DNA was used for PCR and capillary electrophoresis as described previously.

3. Results

3.1 Sensitivity cell separation

In order to evaluate the sensitivity of the cell separation technique, cell suspensions with an increasing ratio of SC/EC were analysed in triplicate using both methods. The average percentage of the detected male alleles was calculated to determine the success rate of STR profiling (Figure 1). Informative male DNA profiles (*i.e.*, alleles of male origin called in ≥10 autosomal loci, excluding alleles shared with the female donor) could be obtained up to a ratio of 1/1000 and 1/5000 using DL and FACS, respectively. Single male DNA profiles could consistently be produced following FACS up to a ratio of 1/100 SC/EC, after which the risk of female contamination of the sperm fraction increased with higher ratios. In contrast, no single male DNA profiles could be produced following DL.

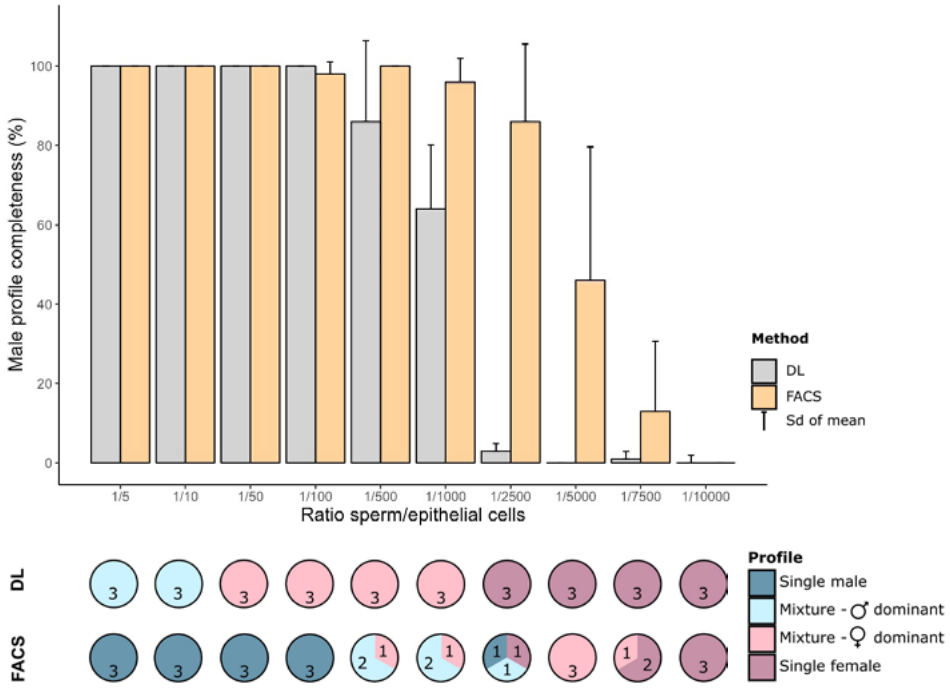


Figure 1. Average male STR profile completeness of samples constituted of 500,000 EC and a declining number of SC. Each ratio was analysed in triplicate using both methods. The pie charts below the bar chart indicate whether the acquired DNA profiles were of single or mixed origin for each respective ratio. Note that alleles shared with the female donor were excluded from these analyses to avoid bias in mixed profiles.

3.2 Mock casework samples – post-coital swabs

To assess the impact of the interval between intercourse and sampling on the efficiency of both methods to produce male DNA profiles, PC swabs were self-collected in duplicate by volunteers (n = 6 couples) at different predetermined timepoints. Complete male DNA profiles were obtained at all timepoints using FACS, whereas allele and locus dropouts increasingly occurred after 48 hours using DL (Figure 2). All DNA profiles obtained using FACS were of single male origin or with minor contamination of the female donor. Conversely, single male profiles could only be acquired with DL within one hour following intercourse.

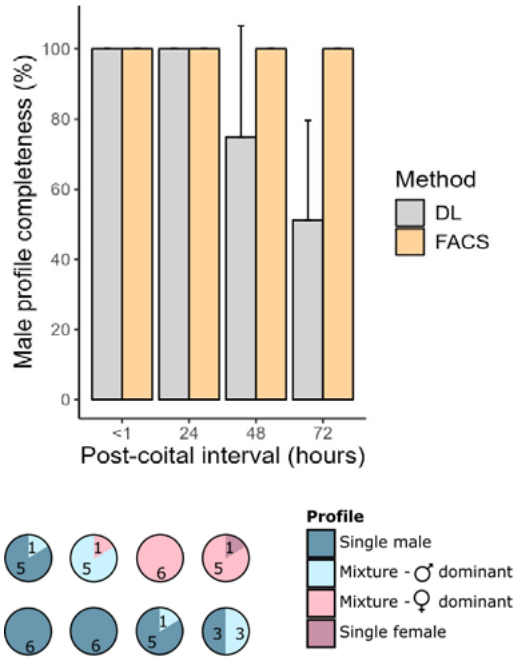


Figure 2. Average male STR profile completeness of PC swabs self-collected by volunteers (n = 6 couples). The pie charts below the bar chart indicate whether the acquired DNA profiles were of single or mixed origin of each timepoint respectively. Note that alleles shared with the female donor were excluded from these analyses to avoid bias in mixed profiles.

Discussion

Male DNA profiles were acquired at a higher SC/EC ratio with FACS than with DL. In addition, less contamination of the male fraction by DNA of the female donor was observed. Although these results are favourable, it has previously been demonstrated that techniques that can process self-assembled samples do not necessarily perform well with true PC samples [6]. Analysing PC swabs was therefore considered an essential component in this study, confirming the improved separation efficiency of FACS. Although figures may vary between different laboratories, approximately 20% of vaginal samples are collected after 24 hours following an alleged offence [7,8]. As this interval also coincides with the timepoint after which DL becomes less efficient, the proposed sperm sorting method would be especially

beneficial in such samples. Since full male DNA profiles could still be acquired at 72 hours following intercourse, samples collected at timepoints exceeding this interval will be tested in future research.

Conclusion

Both with the assembled mixtures and PC samples, sperm sorting outperformed DL in terms of male profile completeness and contamination of the male fraction by female DNA. These findings form a firm foundation for the further validation of the cell sorting method according to the ENFSI guidelines [9] in order to test casework samples.

Acknowledgments

The authors acknowledge the KU Leuven Flow and Mass Cytometry core facility for cell sorting support. This research was funded by a KU Leuven grant no. C3/22/046 – µForensiCell. First author K. Fokias was supported by a Fonds Wetenschappelijk Onderzoek (FWO) grant no. 1S38623N.

Conflict of interest statement

The authors declare no conflict of interest.

References

[1] S. Vuichard, U. Borer, M. Bottinelli, C. Cossu, N. Malik, V. Meier, C. Gehrig, A. Sulzer, M.-L. Morerod, V. Castella, Differential DNA extraction of challenging simulated sexual-assault samples: a Swiss collaborative study, (2011). https://doi.org/10.1186/2041-2223-2-11.

[2] H.R. Dash, Advancements in differentiation between sperm cells and epithelial cells for efficient forensic DNA analysis in sexual assault cases, Int J Legal Med (123AD). https://doi.org/10.1007/s00414-024-03285-1.

[3] W.M.J. Schoell, M. Klintschar, R. Mirhashemi, D. Strunk, A. Giuliani, G. Bogensberger, B. Pertl, Separation of Sperm and Vaginal Cells Based on Ploidy, MHC Class I-, CD45-, and Cytokeratin Expression for Enhancement of DNA Typing After Sexual Assault, (1999). https://doi.org/10.1002/(SICI)1097-0320(19990801)36:4.

[4] A. Nakamura, T. Shinozuka, A. Sonoda, K. Kurosaki, K. Saito, A. Takada, Separation of sperm from mixed samples using ultra-high-speed cell sorter followed by DNA typing, 45 45 法科学技術, 19 (2014) 45–55.

[5] K. Fokias, B. Bekaert, Separation of sperm and epithelial cells based on fluorescence-activated cell sorting, Forensic Sci Int Genet Suppl Ser 8 (2022) 239–241. https://doi.org/10.1016/J.FSIGSS.2022.10.048.

[6] A.J. Hopwood, K. Elliott, Forensic DNA research: keeping it real, Int J Legal Med 126 (2012) 343–344. https://doi.org/10.1007/S00414-011-0636-8.

[7] A.E. Fonneløp, H. Johannessen, G. Heen, K. Molland, P. Gill, A retrospective study on the transfer, persistence and recovery of sperm and epithelial cells in samples collected in sexual assault casework, Forensic Sci Int Genet 43 (2019). https://doi.org/10.1016/J.FSIGEN.2019.102153.

[8] D.G. Casey, K. Domijan, S. MacNeill, D. Rizet, D. O'Connell, J. Ryan, The Persistence of Sperm and the Development of Time Since Intercourse (TSI) Guidelines in Sexual Assault Cases at Forensic Science Ireland, Dublin, Ireland, J Forensic Sci 62 (2017) 585–592. https://doi.org/10.1111/1556-4029.13325.

[9] Guideline for internal validation / verification of various aspects of the dna profiling process, (2023). https://doi.org/10.05.2023.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 411-416
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Test of the Hy-liter system for the identification of spermatozoa in sexual assault cases

Frédéric Grosjean[1,*] and Vincent Castella[1]

[1] *Unit of Forensic Genetics, University Center of Legal Medicine, Lausanne - Geneva, Lausanne University Hospital and University of Lausanne, Chemin de la Vulliette 4, 1000 Lausanne 25, Switzerland*

** Corresponding author address: Unit of Forensic Genetics, University Center of Legal Medicine, Lausanne - Geneva, Lausanne University Hospital and University of Lausanne, Chemin de la Vulliette 4, 1000 Lausanne 25, Switzerland. Email : Frederic.Grosjean@gmail.com*

Abstract

The detection of spermatozoa in forensic samples, such as vaginal swabs, provides critical information regarding the presence of sperm. Immunochromatographic tests, like PSA, are prone to both false positives and false negatives. In contrast, the visual observation of multiple spermatozoa strongly indicates the presence of semen. In this study, we compared the fluorescent detection of spermatozoa using the Hy-Liter system with the Christmas tree staining routinely employed in our laboratory. Sets of mock and real samples were used to compare these methods in terms of ease of staining, time required to prepare and analyze entire slides, specificity of spermatozoa staining in mixed biological materials, reproducibility, as well as reagent handling and cost. Based on these comparisons, Christmas tree staining was retained in our laboratory as the preferred method for the time being.

Keywords

Christmas Tree staining, antibody, fluorescence

Introduction

The Sperm Hy-Liter system by Independent Forensics (IFI) utilizes fluorescent antibodies to detect spermatozoa [1,2]. Theoretically, it should facilitate spermatozoa identification compared to the Christmas Tree staining method [3] routinely used in our laboratory. We compared these two methods in terms of ease of staining,

time required to prepare and analyze entire slides, specificity of spermatozoa staining in mixed biological materials, reproducibility, as well as reagent handling and cost, with the goal of potentially adopting this method.

Material and Methods

Gynecological swabs (4N6FLOQSwabs, COPAN) from volunteers were loaded with 20 µL of sperm dilutions (1:20, 1:100, or 1:1000) or mixtures of 20 µL (1:100) sperm with 20 µL of undiluted blood, 20 µL of undiluted saliva, or in the presence of bacteria (Lactobacilli) or yeast (*Candida albicans*, ATCC 10231 2611). Volunteers who provided biological samples signed informed consent forms. Sample recovery, fixation, and staining were carried out as described in IFI's protocol, but with the Universal Buffer (RSID™), which is used for presumptive tests in our lab [4], to resuspend cells from the swabs instead of PBS. Preliminary tests indicated that the performance of the Hy-Liter system was comparable with both buffers. IFI-supplied slides and SuperFrost slides (Epredia) were tested for material fixation and staining. Fluorescence Axioscope 5 (Zeiss) and Olympus BX41 microscopes were used to visualize the Hy-Liter and Christmas Tree slides, respectively.

Results and discussions

During our initial tests with the Hy-Liter system, cells were lost from the IFI-supplied slides during the wash step following fixation, regardless of the buffer (PBS or Universal Buffer) used for material resuspension (Figure 1). This issue persisted even after allowing the material to dry for 2 hours. The problem was resolved by using SuperFrost slides, which feature a positively charged poly-L-lysine coating.

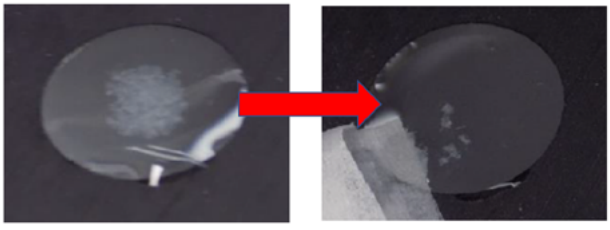


Figure 1. The majority of the material did not adhere to the IFI-supplied slides and was lost during the wash steps. Images of the same slide before (left) and after (right) the wash step (red arrow) are shown. No magnification.

The entire recovery and staining procedure for a single sample was significantly longer with the Hy-Liter system (~ 60 minutes) compared to Christmas Tree staining (~ 20 minutes). Slide screening time was similar for both methods (~ 60 minutes for 6 slides). The reading time with the Hy-Liter system can likely be reduced with increased practice.

The selectivity of the Hy-Liter system was good, with only the spermatozoa being labeled in the presence of the other cell types tested (see the list in the Methods), including yeast and bacteria (Figure 2). As long as the cell density did prevent the formation of aggregates, the Christmas Tree approach performed as well as the Hy-Liter system.

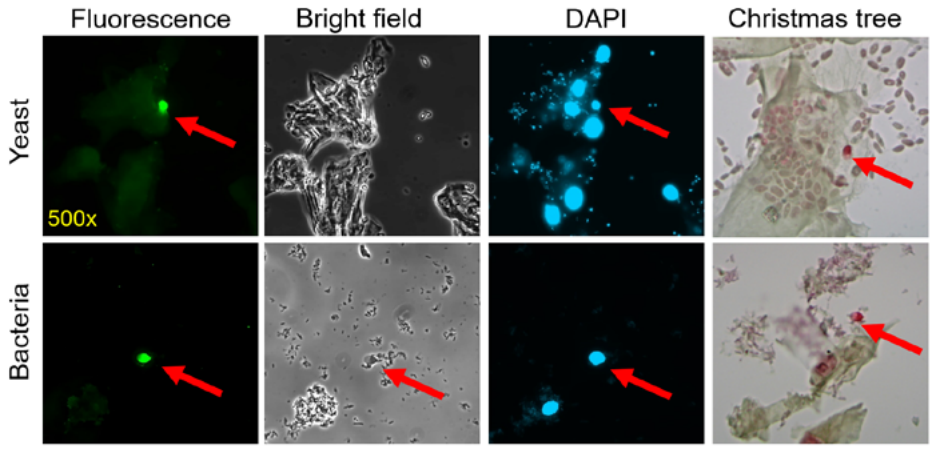


Figure 2. Hy-Liter and Christmas Tree detection of spermatozoa (red arrows) on gynecological swabs loaded with 1:100 diluted semen in the presence of yeast (upper images) or bacteria (lower images). Fluorescent channels are indicated for antibody and DAPI detection, as well as bright field. All images are at the same magnification (as indicated).

Using the Hy-Liter system, variations in staining intensities were observed, as well as ring-staining effects (Figure 3.A) and inhomogeneous staining (Figure 3.B), making the detection of spermatozoa more difficult than expected. We hypothesized that either the relative quantity of DTT used and/or the quantity of specific antibody per spermatozoon was too low. We were unable to find a DTT concentration that would yield similar results in terms of fluorescent intensity between slides prepared from samples with similar characteristics. Non-homogeneous staining was also observed between the center of the staining zone and its periphery for most of the samples when there were large numbers of epithelial cells.

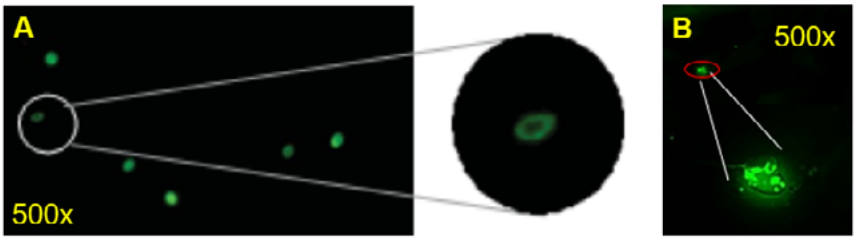


Figure 3. Ring-staining effects (A) of spermatozoa observed with the Hy-Liter system. Inhomogeneous staining of a spermatozoon (B).

The detection of spermatozoa within cellular aggregates is facilitated by fluorescence [1,2]; however, it is not possible to confirm morphologically in the bright field that these are indeed spermatozoa. The same visualization is virtually impossible within aggregates using the Christmas Tree staining (Figure 4.A). Moreover, the autofluorescence of some substrates (for instance, a bath towel) makes the identification of a possible spermatozoon very difficult (Figure 4.B) with the Hy-Liter system.

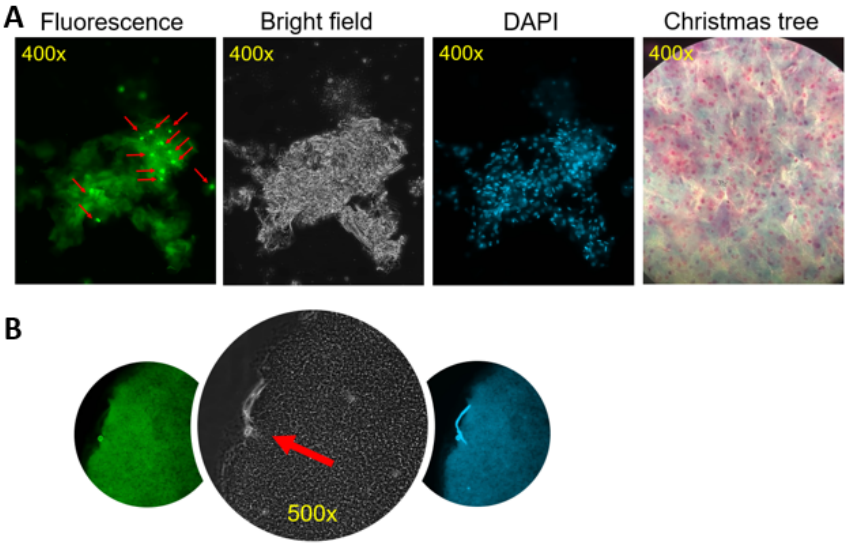


Figure 4. A. Possible spermatozoa detection (red arrows) with the Hy-Liter system (first three images) and absence of such detection with the Christmas Tree staining (fourth image), both in the presence of cellular aggregates (1:100 diluted semen on gynecological swabs). All images

are at the same magnification (as indicated). B. A bath towel where the presence of semen is suspected. Detection is made difficult by the strong autofluorescence of the towel, both in the antibody (left image) and DAPI (right image) fluorescence channels. It's difficult to say whether a spermatozoon is present or not using bright field (middle image, red arrow).

Conclusion

Despite a good selectivity, we haven't retained the Hy-Liter in its current form in our laboratory. This is due to the non-uniformity of the staining, the toxicity of one of the reagents (DAPI), the difficulties in confirming the presence of spermatozoa in some samples, the difference in reagent costs per sample (Hy-Liter ≈ 20-25 €; Christmas Tree ≈ 0.5 €), and the lack of all-in-one conditions that would work with all sample types.

Acknowledgments

The authors wish to thank all donors who took part in this experiment, especially Iris Schulz and Gunn-Hege Olsen for sharing their knowledge with us regarding the Hy-Liter system.

Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this article.

References

[1] A. De Moors et al, Sperm Hy LiterTM: an effective tool for the detection of spermatozoa in sexual assault exhibits, Forensic Sci. Genet. 7 (2013) 367–379. PMID: **23571070 /** doi.org/10.1016/j.fsigen.2013.02.011

[2] K.W. Miller et al, Developmental validation of the sperm HY-LITER kit for the identification of human spermatozoa in forensic samples, J. Forensic Sci. 56 (2011) 853–865. PMID: **21595688 /** doi.org/10.1111/j.1556-4029.2011.01796.x

[3] J.P. Allery et al, Cytological detection of spermatozoa: comparison of three staining methods, J. Forensic Sci. 46(2) (2001) 349-351. PMID: 11305439

[4] P. Basset et al, A simplified protocol for the detection of blood, saliva, and semen from a single biological trace using immunochromatographic tests, Forensic Sci Med Pathol 18(2) (2022) 141-148. PMID: 35171453 / doi.org/10.1007/s12024-021-00453-2

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 417-422
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Development of a new automated EZ2® Connect Fx sexual assault sample processing kit

Sarah Aurora Hess[1], Melanie Kraemer[1], Daniele Terramagra[1], Nina Golob[1], Philipp Rink[1], Stefan Cornelius[1], Anke Prochnow[1], Amy Liberty[2], Marina Nastainczyk-Wulf[3], Mackenzie DeHaan[4], Erin Ermish[4], Shane Wilcox[4], Jody West[4], Ryan O'Leary[5], Dustin Bell[5], Paul Berry[5], Mario Scherer[1]

[1]*QIAGEN GmbH, Hilden, Germany.*
[2]*QIAGEN LLC, Germantown, USA.*
[3]*Institute of Legal Medicine, University Hospital Halle (Saale), Germany.*
[4]*North Carolina Department of Justice, Raleigh, USA.*
[5]*Louisiana State Police Crime Laboratory, Baton Rouge, USA.*

** Corresponding author: Sarah Aurora Hess, QIAGEN GmbH, QIAGEN Straße 1, 40724 Hilden, Deutschland, Phone +4921032914128, sarah.hess@qiagen.com*

Abstract

A new sexual assault sample processing kit has been developed to address the need for fast processing and generation of distinct male profiles, which is often insufficient with traditional routine methods. **The new kit enhances sperm DNA recovery and separation from mixed samples. It operates on the EZ2 Connect Fx platform, offering improved throughput and user-friendliness, with just 30 minutes of hands-on time for up to 24 samples.**

The method delivers fast, reproducible results, addressing the need for timely analysis and backlog reduction. Sperm fractions are ready-to-use without further purification and compatible with common kits for downstream applications like qPCR and STR profiling on CE and NGS.

Particularly effective for challenging samples, like 66-hour post-coital swabs, the prototype has shown significant improvements over existing methods. Alpha-field tests conducted across four laboratories revealed that the prototype generated sperm fractions with fewer alleles that originated from the female donor, compared to the QIAcube workflow. The prototype also recovered a higher number of male alleles, often with higher RFUs than those from the QIAcube method. This effective separation of male DNA from mixed samples resulted in single-person STR profiles or mixed profiles with distinct male components, with no need for probabilistic genotyping software for interpretation.

In conclusion, the EZ2 Connect Fx prototype offers significant advancements in forensic DNA analysis, especially for complex cases, enabling faster, more precise processing of sexual assault evidence.

Keywords
Sexual assault, semen, sperm, spermatozoa, rape.

1. Introduction

Analysis of sexual assault samples often faces challenges due to insufficient separation of genetic information from the victim and perpetrator, resulting in mixed profiles. A new sample processing kit has been developed to efficiently lyse non-sperm cells (e.g., epithelial cells) while improving sperm DNA recovery and precise separation between sperm-derived DNA and DNA coming from non-sperm cells.

The kit offers automated processing on the EZ2 Connect Fx with increased throughput and ease of use. With just 30 minutes of hands-on time for 24 samples, it delivers fast, reproducible results, addressing the need for timely analysis and backlog reduction in sexual assault cases.

Data from current development of the new EZ2 sexual assault sample processing prototype kit is shown, as well as alpha-field test results generated by four laboratories, using various sample types.

2. Material studied, methods, techniques

2.1. Samples

A range of sexual assault samples was used (Tab. 1). Vaginal swabs spiked with 1:50 and 1:400 semen dilutions were used in all four laboratories. Additionally, three external forensic laboratories tested a wide range of specific samples.

Table 1. Overview of samples used for alpha-field test in all four laboratories.

	QIAGEN + NCDOJ + LSPCL + ILM Halle		NCDOJ (North Carolina, USA)			LSPCL (Louisiana, USA)			ILM (Halle, Germany)		
Sample	1:50	1:400	PC 24 h	PC 36 h	Sheet	1:150	1:400	1:600	PC 10 h	PC 66 h	Oral wash
Sample type	Vaginal swabs, spiked with semen dilutions (1:50; 1:400)		24 h post-coital (7 years)	36 h post-coital (7 years)	Used sheet + semen stains	Buccal swabs, spiked semen dilutions: 1:150 (<2 years), 1:400 (10 years), 1:600 (4 years) (3 months) (1 month)			10 h post-coital	66 h post-coital	Mouth-wash + semen
STR kit	24plex QS		PowerPlex Fusion 6C; Y23			PowerPlex Fusion 6C			ESX17 Fast		

2.2. Methods

Samples were processed using the automated EZ2 Connect Fx prototype kit and compared to the differential wash workflow on the QIAcube® instrument. Sample preparation for the prototype involved an off-deck process (1 hour, 10 min) including lysis, centrifugation with a spin basket, removal of solid components and loading onto the EZ2 Connect Fx. The automated process (1 hour, 5 min) removed the epithelial fraction (EPI), separated and purified the sperm fraction in three steps. After sperm cell lysis, an inactivation step (15 min) was performed. This method processed up to 24 sperm fraction eluates in 2.5 hours. The epithelial fraction was purified using the EZ1&2 DNA Investigator Kit or similar methods.

The final kit will be fully automated on the EZ2 Connect Fx with a run time of 1 hour, 10 minutes, including inactivation.

DNA quantification and STR profile generation followed standard laboratory procedures (Tab. 1). An analytical threshold of 200 RFU was used for STR interpretation (except for NCDOJ samples: 50 RFU), with an internally used locus-specific threshold of ≥9% RFU of the highest allele within the STR locus. For concordance studies, reference STR profiles were used, where each allele of the STR profiles was assigned to the correct donor. No reference profile was available for sample 'Sheet' (NCDOJ).

3. Results & Discussion

3.1. Reduced number of STR alleles of the female donor of the sperm fraction

STR results from sperm fractions generated with the prototype showed significantly fewer female donor alleles compared to the QIAcube workflow (Fig. 1). The prototype demonstrated good reproducibility across various mixed samples. For challenging samples, like 10 µL of 1:400 diluted semen on vaginal swabs or 66-hour post-coital swabs, the prototype detected more alleles that originated from the female donor compared to routine samples. However, even with these challenging samples, it exhibited a significantly reduced number of alleles that originated from the female donor, compared to the QIAcube method, resulting in male single-person profiles or mixed profiles with a distinct male major component.

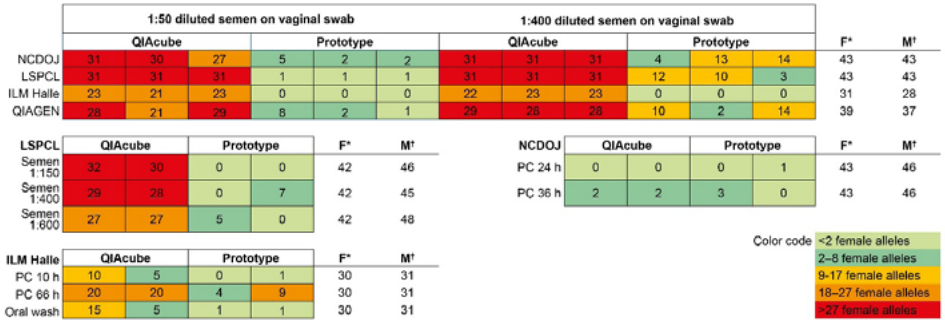


Fig. 1 Comparison of female allele counts of the sperm fraction between different methods and laboratories. Using reference STR profiles, each allele of the STR profiles was assigned to the correct donor. The number inside the cells indicates the number of STR alleles of the female donor. Each cell is one biological replicate. *Number of female donor alleles in the STR kit; †Number of male donor alleles in the STR kit.

3.2. Precise separation and distinct profiles even with challenging samples

The prototype produced profiles with fewer alleles and significantly lower RFUs of alleles that originated from the female donor, while male donor alleles showed higher RFUs, compared to the QIAcube differential wash. This was particularly effective for challenging samples with high DNA of the female donor and low sperm quantities (Fig. 2A).

The prototype accurately separated sperm DNA from mixed samples, yielding male single-person profiles or mixed profiles with a distinct male component. In contrast, the QIAcube method worked well with high male DNA but showed mixed profiles with challenging samples.

Sarah Aurora Hess, Melanie Kraemer, Daniele Terramagra, Nina Golob, Philipp Rink, Stefan Cornelius, Anke Prochnow, Amy Liberty, Marina Nastainczyk-Wulf, Mackenzie DeHaan, Erin Ermish, Shane Wilcox, Jody West, Ryan O'leary, Dustin Bell, Paul Berry, Mario Scherer

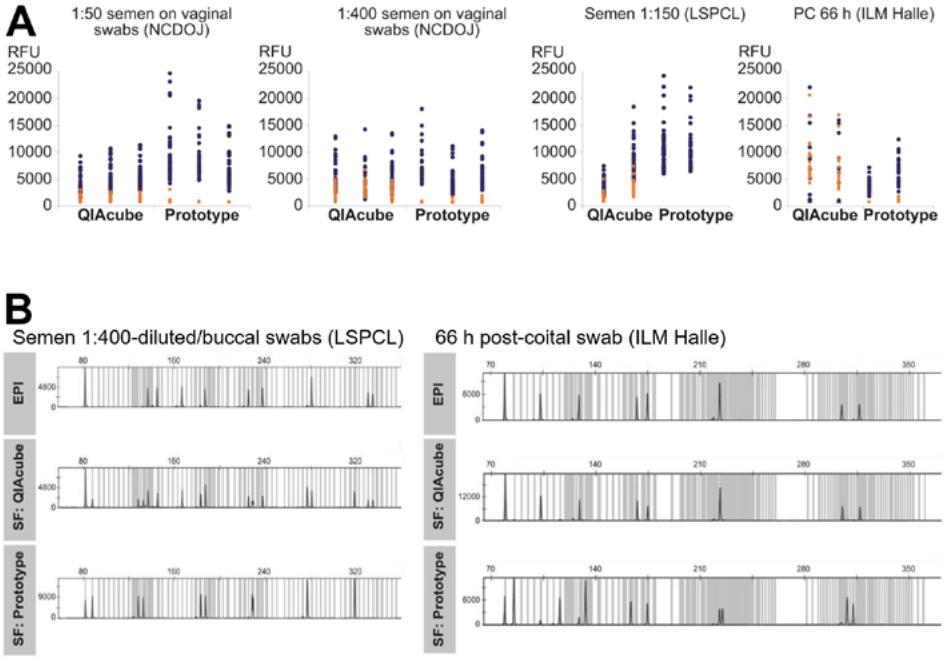


Fig. 2 Representative examples of the comparison between the QIAcube method and the prototype method. (A) Using the respective reference profiles of the DNA donors, each allele of the STR profiles was assigned to the correct donor and highlighted according to origin: Female donor (orange dot), or male donor including shared alleles (blue dot). One strip = one sample, i.e., one specific sperm fraction STR profile (coming from one biological replicate). **(B)** Representative parts of electropherograms (one channel). All profiles of the sperm fractions (SF) were tested against the reference profiles of female and male donors. The epithelial fraction (EPI) always shows the profile of the female donor.

The effective separation of distinct male STR profiles from mixed samples is demonstrated in two examples (Fig. 2B). In a mock sample with 10 µL of 1:400 diluted semen on buccal swabs (left), the QIAcube produced a mixed profile, while the prototype generated a single-person male profile from the same sample. For 66-hour post-coital swabs (right), both methods produced mixed profiles. However, the QIAcube showed a female major component, while the prototype yielded a distinct male major profile. No probabilistic genotyping software was needed to interpret the prototype results in either case.

4. Conclusion

The new automated EZ2 Connect Fx sexual assault sample processing prototype kit demonstrated enhanced precision and reliability, especially in complex scenarios where traditional methods struggle to provide distinct profiles.

The prototype kit improves separation between sperm-derived DNA and non-sperm DNA, enabling efficient processing of various sexual assault sample types. It is particularly beneficial for challenging cases, such as post-coital samples that show low sperm quantity and high victim DNA amount. Proven performance in benchmarking experiments across four laboratories has shown high reproducibility across a wide range of mixed samples. Key advantages include reduced hands-on time, ease-of-use, and increased throughput, allowing faster processing and addressing large backlogs. Additionally, the workflow enables direct use of sperm fractions without further purification and is fully compatible with commercial kits such as qPCR, STR, and NGS.

5. Acknowledgments

None.

6. Conflict of interest statement

None.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 423-428
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Alkaline lysis of sperm cells from forensic sexual assault samples

Christina Forsberg[1,2*], Elin Lundkvist[1], Johannes Hedman[1,2]

[1]*National Forensic Centre, Swedish Police Authority, Linköping, Sweden*
[2]*Applied Microbiology, Department of Chemistry, Lund University, Lund, Sweden*

** Corresponding author at: National Forensic Centre, Swedish Police Authority, Linköping, Sweden.*
E-mail: christina-u.forsberg@polisen.se Telephone: 0046105628131

Abstract

Differential extraction is typically used for forensic sexual assault samples in order to separate sperm cells (from the perpetrator) from other cells (predominantly from the victim). Sperm cells are more robust compared to e.g. white blood cells and epithelial cells and therefore require more stringent lysis conditions. Traditionally, the reducing agent DTT in combination with Proteinase K is used in the digestion of sperm cells in forensic samples. Since DTT may inhibit the downstream PCR process, the extracts must be purified. Purification may be performed using e.g. phenol/chloroform, magnetic bead-based chemistries or filter devices. Application of any of these methods increases time and cost and leads to DNA loss. In this study, we evaluated alkaline lysis as an alternative to the traditional lysis using DTT and Proteinase K. In the alkaline lysis protocol, NaOH is used to disrupt the cell and nucleus membranes and denature nucleases. This is followed by the addition of Tris-HCl to neutralize the lysate. The effect on DNA yield of different molar concentrations of NaOH and Tris-HCl was evaluated to determine the optimal lysis conditions. Overall, the chosen alkaline protocol showed a higher DNA recovery compared to lysis using DTT/Proteinase K (mean values of 0.075 vs. 0.035 ng/µL, p=0.025). The quality of the STR profiling results was equivalent for the two lysis principles. STR typing showed no indication of degradation after storage in the refrigerator or freezer for up to two weeks. In addition, there was no need for a purification step following alkaline sperm cell lysis, reducing the cost, hands-on time and total processing time.

Keywords

Alkaline lysis, DNA extraction, differential lysis, forensic sexual assault samples.

Introduction

Differential extraction is used for forensic sexual assault samples in order to separate sperm cells (typically from the perpetrator) from other cells (predominantly from the victim). Sperm cells are more robust compared to e.g. white blood cells and epithelial cells and therefore require more stringent lysis conditions. Traditionally, the reducing agent dithiothreitol (DTT) in combination with Proteinase K is used in the digestion of sperm cells in forensic samples. The DNA in the sperm nucleus is mainly bond to protamines. Covalent disulfide bonds between the protamines allow the DNA to be tightly packed and protected against damage (Figure 1) [1,2]. DTT is used to break the disulfide bonds to make the DNA available for PCR. Since DTT may inhibit the downstream PCR process, the extracts must be purified. Purification may be performed using e.g. phenol/chloroform, magnetic bead-based chemistries or filter devices. Application of any of these methods increases time and cost and leads to DNA loss.

In order to circumvent time-consuming incubation and purification steps we investigated an alternative lysis of sperm cells by an alkaline lysis protocol. There, NaOH is used to disrupt the cell and nucleus membranes and denature nucleases [3,4], followed by the addition of Tris-HCl to neutralize the lysate. The effect on DNA yield of different concentrations of NaOH and Tris-HCl was evaluated to determine the optimal lysis conditions.

Material and methods

A semen sample (provided by a volunteer donor) was diluted with physiological saline and a volume corresponding to 0.05 µL of semen was added to microfuge tubes. All samples were extracted using the epithelial cell lysis steps of an optimized differential lysis protocol based on reference [5]. Alkaline lysis was then performed by adding 10 or 25 µL of 0.1 M, 0.2 M or 0.5 M NaOH (Merck Sigma-Aldrich, Darmstadt, Germany) to the sperm pellet. After incubation in a thermomixer at 75°C for 5 minutes an equivalent amount of Tris-HCl (10 or 25 µL of 0.1 M, 0.2 M or 0.5 M Tris-HCl, Thermo Fisher Scientific, Waltham, MA, USA) was added. TE buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0, Medicago AB, Uppsala, Sweden) was used to reach a final volume of 100 µL. The stability of the DNA extracts was evaluated by keeping the samples in the refrigerator or freezer for up to 14 days.

As reference method, sperm pellets were also extracted by adding 170 µL of 5% Chelex in Milli-Q water (Bio-Rad Laboratories, Hercules, CA, USA), 40 mM DTT and 0.4 mg/mL Proteinase K (Merck Sigma-Aldrich). After incubation in a heating incubator (56°C for 30 minutes and 100°C for 20 minutes) the extracts were purified using Amicon Ultra-2 filter devices (Merck Millipore, Darmstadt, Germany) and TE buffer, with a final volume of 100 µL.

The samples were quantified with PowerQuant System (Promega Corporation, Madison, WI, USA) on a 7500 real-time PCR system (Thermo Fisher Scientific), amplified with the STR kit PowerPlex Fusion 6C System (Promega) on a Veriti Thermal Cycler, injected on a 3500xL Genetic Analyzer and analysed using GeneMapper ID-X v.1.6 (Thermo Fisher Scientific).

Results and discussion

All tested concentrations and volumes of NaOH gave higher DNA yields compared to the reference method ($p<0.05$) as shown in Figure 2. The PowerQuant Internal PCR Control did not show any signs of inhibition. The quality of the STR profiling results was equivalent for the two lysis principles (DTT and Proteinase K vs. alkaline lysis) when using 0.1 M (10 and 25 µL) or 0.2 M (10 µL) NaOH (apart from a few replicates with NaOH showing slightly lower peak heights for Amelogenin). The addition of higher concentrations of NaOH lead to impaired amplification for the longer STR markers (data not shown). Thus, 10 µL of 0.2 M NaOH was chosen for further testing.

STR typing showed no indication of DNA degradation after storage in the refrigerator or freezer for up to 14 days. After keeping the samples in the refrigerator for 14 days the mean DNA concentration was somewhat lowered (0.062 ng/µL vs. 0.046 ng/µL), but the difference was not significant ($p=0.08$) (Table 1).

Conclusion

The chosen alkaline protocol showed a higher DNA recovery compared to lysis using DTT/Proteinase K (mean DNA concentrations 0.075 vs. 0.035 ng/µL, $p=0.025$). STR typing showed no indication of DNA degradation after storage in the refrigerator or freezer for up to 14 days. There was no need for a purification step when using alkaline sperm cell lysis, thereby reducing the cost, hands-on time and the total processing time.

Conflict of interest statement

The authors declare no conflicts of interest.

References

[1] D. Mortimer, The functional anatomy of the human spermatozoon: relating ultrastructure and function, Molecular Human Reproduction 24:12 (2018) 567-592.

[2] W.S. Ward and D.S. Coffey, DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells, Biology of Reproduction 44:4 (1991) 569-74.

[3] W.R. Hudlow and M.R. Buoncristiani, Development of a rapid, 96-well alkaline based differential DNA extraction method for sexual assault evidence, Forensic Science International: Genetics 6:1 (2012) 1-16.

[4] S.K. Schellhammer, B.C. Hudson, J.O Cox, et al., Alternative direct-to-amplification sperm cell lysis techniques for sexual assault sample processing, Journal of Forensic Sciences 67:4 (2022) 1668–1678.

[5] AmpFlSTR® Profiler PCR Amplification Kit, User's Manual. Part Number 402945 Rev. G, 03/2012.

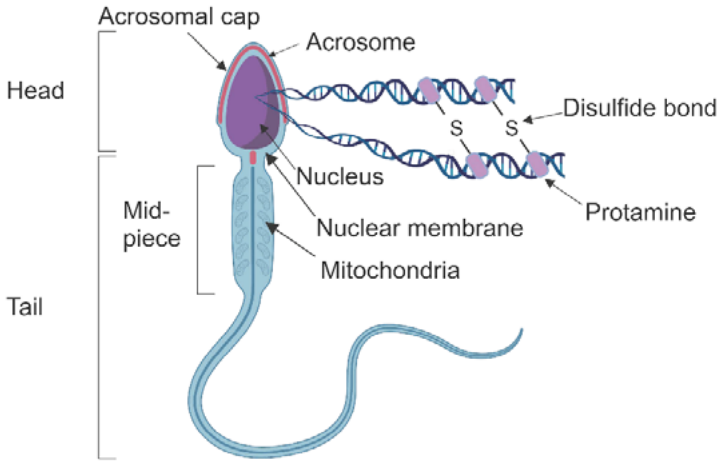


Figure 1. Spermatozoon cell structure (created with BioRender.com).

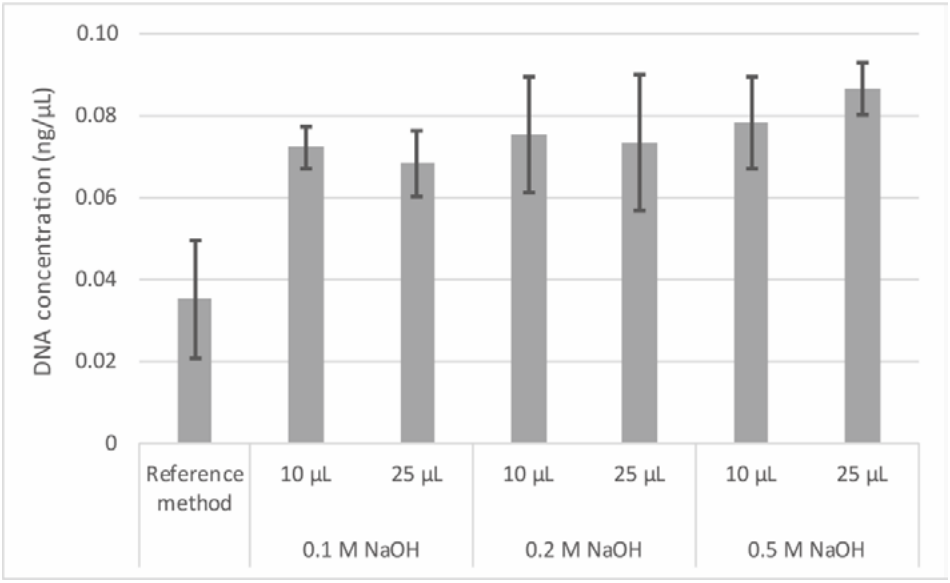


Figure 2. Mean DNA concentrations (ng/µL) when adding different molarities and volumes of NaOH to the sperm fractions compared with the reference method, n=3. Error bars show standard deviation.

Table 1. Mean DNA concentrations (ng/µL) and coefficient of variation (CV, %) directly after DNA extraction and after keeping the extracts from alkaline lysis (10 µL of 0.2 M NaOH) in freezer or refrigerator for 7 or 14 days, n=5.

Day	Storage	DNA concentration (ng/µL)	CV (%)
0	-	0.062	13
7	Refrigerator	0.062	16
7	Freezer	0.062	16
14	Refrigerator	0.046	30
14	Freezer	0.064	19

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 429-435
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Usefulness of one-off DNA database searches in Western Switzerland

Patrick Basset[1*], Lydie Samie[1], Frédéric Grosjean[1] and Vincent Castella[1]

[1] Unit of Forensic Genetics, University Center of Legal Medicine, Lausanne - Geneva, Lausanne University Hospital and University of Lausanne, Chemin de la Vulliette 4, 1000 Lausanne 25, Switzerland

** Corresponding author address: Unit of Forensic Genetics, University Center of Legal Medicine, Lausanne - Geneva, Lausanne University Hospital and University of Lausanne, Chemin de la Vulliette 4, 1000 Lausanne 25, Switzerland. Email : Patrick.Basset@chuv.ch*

Abstract

DNA profiles from biological traces that are uploaded to forensic DNA databases (DNADB) must generally meet specified criteria to lower the risk of adventitious associations. For example, in Switzerland, only single-source DNA profiles with a minimum of 6 loci and two-person mixtures with a minimum of 8 loci can be transmitted to the national DNADB. Interpretable DNA profiles that do not meet these criteria represent about 14% of the ~15'000 trace DNA profiles analyzed each year in Western Switzerland. These profiles can be manually compared with candidates upon request from police services. Alternatively, they can be compared once within the national DNADB without being stored (i.e., one-off search). In this study, we reviewed the one-off searches performed in Western Switzerland between 2014 and 2023. The number of requests for one-off searches increased significantly during this period, particularly since the introduction of probabilistic genotyping in our laboratory in 2019. Since then, 66 compatibilities with LR > 1 were observed out of 298 requests, representing investigation leads for more than 20% of the requests. The usefulness of one-off searches, as well as their limitations, are discussed.

Keywords

DNA database search, probabilistic genotyping, complex DNA profile

1. Introduction

Forensic DNA databases (DNADB) are effective **in providing** investigation leads (1). Generally, DNADBs compare two groups of samples: one **consisting of** DNA profiles recovered from forensic traces and a second **composed of** DNA profiles from known persons. Comparisons between these two groups of profiles can result in the association of DNA profiles from traces with those from their potential sources. The risks of adventitious associations increase with DNA profiles having common alleles, the number of contributors, and the size of the DNADB. Therefore, trace DNA profiles need to **meet** various criteria to be uploaded into a DNADB (2). These criteria vary among countries, but DNA profiles usually must include a minimum number of validated alleles or loci, and a **maximum** number of contributors (3, 4). In Switzerland, for example, only profiles from one contributor (single or major contributor) **involving** at least six validated loci (out of 16), and mixture DNA profiles from a maximum of two contributors **with** at least eight validated loci can be sent to the national DNADB. In contrast, interpretable profiles such as mixtures from more than two contributors, partial DNA profiles, and minor fractions of DNA mixtures cannot be sent to the national DNADB. We refer to these profiles as complex DNA profiles in this article.

Touch DNA samples, also called trace DNA samples, represent a large proportion of crime scene samples currently analyzed. They often contain low quantities of DNA, possibly from several individuals, and resulting DNA profiles are often partial and/or mixed. Thus, many of these profiles cannot be transmitted to a DNADB. For example, about 55% of the ~15,000 traces analyzed in Western Switzerland in 2023 generated interpretable DNA profiles, whereas about 45% generated DNA profiles that couldn't be used because of their poor quality or complexity (i.e., no DNA profile, unreproducible replicates, and/or involving more than five contributors). The 55% of interpretable DNA profiles can be sub-divided into those that met the criteria for uploading into the Swiss national DNADB (i.e., one or two contributors, accounting for 45% of the traces analyzed) and the DNA profiles that couldn't be transmitted to the DNADB, mainly because they involved 3 to 5 contributors (accounting for the remaining 10% of the interpretable traces). In addition, 4% of the DNA profiles uploaded to the national DNADB contain a minor fraction with sufficient information to allow for comparisons with known DNA profiles. These complex DNA profiles can be manually compared with candidates upon request from the police services. Alternatively, these DNA profiles can be compared once (without being stored) with the national DNADB using information selected by

the DNA scientist. This process is called a one-off search (5). The aim of this study was to determine the usefulness and limitations of one-off searches in Western Switzerland.

2. Material and methods

All **one-off search cases** performed in Western Switzerland between 2014 and 2023 were reviewed. This region was chosen because very few one-off searches were requested in other parts of Switzerland. For instance, 107 of the 112 one-off searches performed in Switzerland in 2023 came from Western Switzerland. The situation was similar throughout the study period. One-off searches were considered successful when the DNA profile of at least one candidate was visually considered **"compatible" before 2019**, or when a likelihood ratio (LR) higher than 1 was assigned using STRmix™ (6) for at least one candidate from 2019. The DNA profile of a candidate was considered compatible when its alleles were observed in the **DNA mixture profile** and when the alleles' heights matched expectations. In the case of discrepancies, these had to be explained by possible artifacts such as drop-outs or allele imbalances.

3. Results

One-off searches were carried out for 421 trace DNA profiles analyzed in Western Switzerland between 2014 and 2023. This represents less than 10% of the complex DNA profiles generated during this period. The number of candidates proposed following each one-off search varied between 0 and 1,242 (**mean candidate number** = 6 and **median** = 1). Interestingly, the number of requests increased during the study period, particularly since the introduction of probabilistic genotyping (PG) in our laboratory in 2019 (Figure 1). Over the years, the success rate of these searches varied between 9% and 36%, with a **mean** of 23% (Figure 2).

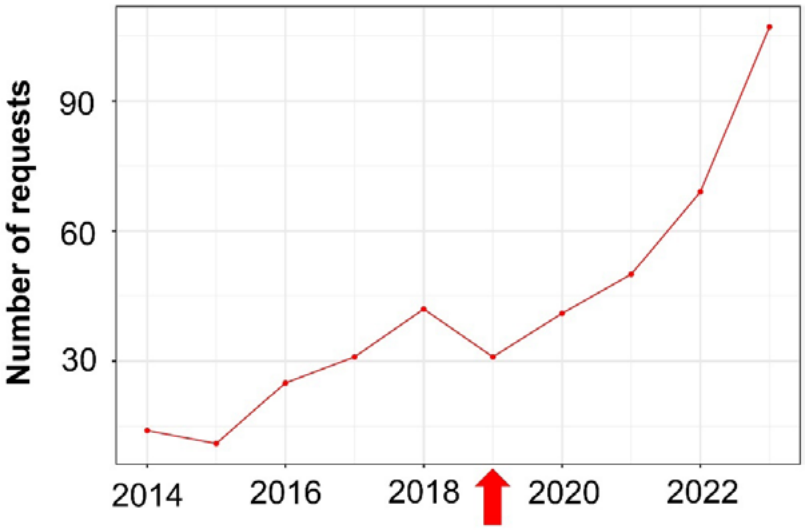


Figure 1. Number of one–off searches performed in Western Switzerland between 2014 and 2023. The red arrow indicates the introduction of probabilistic genotyping in the laboratory.

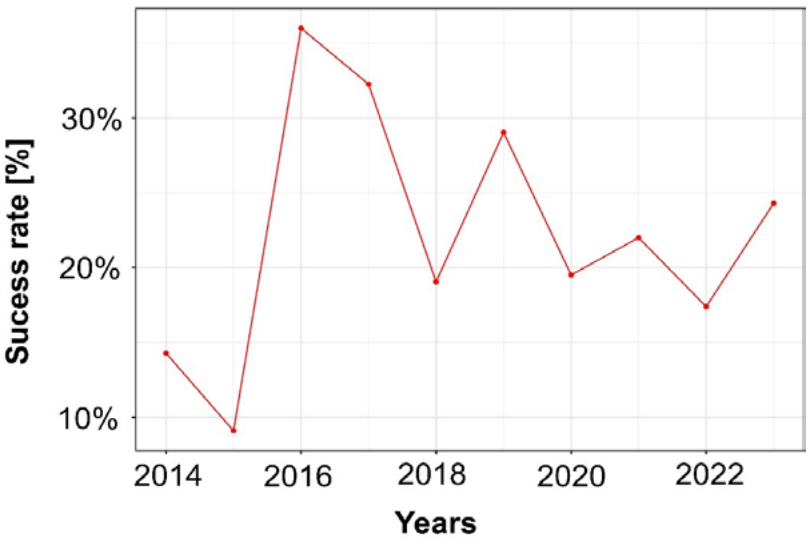


Figure 2. Success rate of one–off searches during the study period defined as visual compatibilities before 2019 and with LR > 1 from 2019.

4. Discussion

Our study showed that one-off searches within the Swiss DNADB are effective, as **compatibilities**, and therefore investigation leads, were obtained in 23% of the cases reviewed. The number of candidates proposed by the searches was generally low, as the median number of candidates per search was one. This number can be easily managed by investigators. Nevertheless, some searches might result in long lists of candidates, for example, up to **1,242** in a given case. Several parameters, such as the size of the DNADB or the information contained in the DNA profile (rarity of the observed alleles, number of contributors, relative quantity of each DNA, etc.), can affect the number of candidates proposed as well as the success rate. Depending on these parameters, the search criteria should be adapted, considering the risk of adventitious links and the risk of missing real contributors. Generally, only part of the information contained in DNA mixture profiles is used for one-off searches, which may result in partial DNA profiles. Often, the number of candidates can be reduced by using all the information contained in the DNA profiles, either visually or using probabilistic genotyping (PG) software (7). PG software can be used to prioritize among candidates based on the likelihood ratio (LR). The benefit of PG is illustrated by the increase in the number of requests since its introduction in our laboratory in 2019. Before PG, DNA profiles from candidates could only be declared as visually compatible/incompatible, without any quantitative information on the weight of this comparison result. Using PG, it is possible to assign a LR to determine the value of the compatibilities. Since 2019 and the introduction of PG in our laboratory, 66 compatibilities with LR > 1 were obtained out of 298 requests. High LRs, in the order of billions, were obtained for DNA profiles with up to 5 contributors. This highlights the value of these investigation leads, even for complex DNA profiles. It is, however, important to consider these compatibilities in conjunction with other information related to the case (8).

Yet, one-off searches have some **limitations.** First, trace DNA profiles used for one-off searches are not stored in the DNADB. Therefore, if the DNA profile of a contributor is introduced into the DNADB after the search, the link will be missed. If necessary, searches can be resubmitted periodically when the DNA profiles of new individuals have been uploaded into the DNADB. Second, the searches are generally performed with only part of the DNA profile to reduce the number of candidates. It is therefore possible to miss an association with the DNA profile of a contributor currently in the DNADB if the corresponding alleles were not used for the search. This means that an absence of association

does not necessarily mean that none of the contributors' profiles are present in the DNADB. This limitation can be partially overcome by comparing the full profile using specialized PG software (9, 10). Finally, assessing candidates is time-consuming, especially in situations with long lists of candidates. This probably explains why one-off searches were requested for less than **10%** of the complex DNA profiles. The selection of DNA profiles for one-off searches should, therefore, be based on a cost-benefit analysis considering the type of DNA profiles involved, the severity of the case, as well as the resources available.

5. Conclusion

Complex DNA profiles, comprising DNA mixtures of 3, 4, and 5 contributors, as well as minor fractions of DNA profiles, represent a non-negligible part of the DNA profiles analyzed. Our data show that these currently underused DNA profiles should not be discarded. Similarly to single source DNA profiles, they can be efficiently compared to databases of potential offenders using one-off searches. Currently, CODIS is used to compare DNA profiles at a national level, and we use PG to assign likelihood ratios for associations involving DNA mixtures. In the future, we hope that complex DNA profiles can be directly compared with DNADB using specialized PG software (9, 10).

5. Acknowledgments

We would like to thank the Swiss DNADB for the information provided about one-off search in the country.

6. Conflict of Interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7. References

[1] Amankwaa AO. Trends in forensic DNA database: transnational exchange of DNA data. Foren Sci Res. 2020;5(1):8-14.

[2] Hicks T, Taroni F, Curran J, Buckleton J, Ribaux O, Castella V. Use of DNA profiles for investigation using a simulated national DNA database: Part I. Partial SGM Plus® profiles. Forensic Sci Int-Gen. 2010;4(4):232-8.

[3] ENFSI DNA Working Group. DNA Database Management Review and Recomandations. https://enfsi.eu/news/guideline-dna-database-management-review-and-recommendations/2019.

[4] Wilson L. Combined DNA Index System: CODIS. In: Houck MM, editor. Encyclopedia of Forensic Sciences, Third Edition (Third Edition). Oxford: Elsevier; 2023. p. 596-604.

[5] Pope S, Puch-Solis R. Interpretation of DNA data within the context of UK forensic science - investigation. Emerg Top Life Sci. 2021;5(3):395-404.

[6] Bright JA, Taylor D, McGovern C, Cooper S, Russell L, Abarno D, et al. Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles. Forensic Sci Int-Gen. 2016;23:226-39.

[7] Coble MD, Bright JA. Probabilistic genotyping software: An overview. Forensic Sci Int-Gen. 2019;38:219-24.

[8] Hicks T, Buckleton J, Castella V, Evett I, Jackson G. A Logical Framework for Forensic DNA Interpretation. Genes-Basel. 2022;13(6).

[9] Bright JA, Taylor D, Curran J, Buckleton J. Searching mixed DNA profiles directly against profile databases. Forensic Sci Int-Gen. 2014;9:102-10.

[10] Benschop CCG, van de Merwe L, de Jong J, Vanvooren V, Kempenaers M, van der Beek CP, et al. Validation of SmartRank: A likelihood ratio software for searching national DNA databases with complex DNA profiles. Forensic Sci Int-Gen. 2017;29:145-53.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 437-441
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Incorporating short tandem repeat sequences into STRBase: Past, Present, and Future

L.A. Borsuk[1*], P.M. Vallone[1], K.B. Gettings[1]

[1]*U.S. National Institute of Standards and Technology, NIST 100 Bureau Drive. Gaithersburg, MD 20899-8314, U.S.A.*

** Corresponding author – lisa.borsuk@nist.gov*

Abstract
STRBase (strbase.nist.gov) was launched in late 1997 to collect and organize information in one location about Short Tandem Repeats (STR), the new emerging forensic marker type at the time. Over the next twenty-plus years, data related to STR markers was collected from the community and incorporated into STRBase. More recently, this has included adding the STR sequencing information from the STRSeq project, which is a curated set of sequences collected from the forensic community housed at the National Center for Biotechnology Information (NCBI) in the form of GenBank records organized into BioProjects. STRBase and STRSeq continue to evolve and grow, with additional functionality and information continuously added.

Keywords
Short Tandem Repeat (STR), Sequence, STRSeq, STRBase

Past

The original STRBase website was developed as a collection of individual web pages. These pages included fact sheets about loci and, if available, the sequence of STR alleles as performed at the National Institute of Standards and Technology (NIST) or reported in the literature by Sanger sequencing methods. The reported information on the site was the bracketed repeat and the length-based allele call. These hard-coded tables on the locus fact sheet webpages can still be viewed on the archived site [1].

In 2017, the development of a STRBase site based on.NET and C# with a backend database began. The result of this effort is the publicly available STRBase site [2]. During the development, the content of the HTML pages was evaluated, and it was determined that the sequence tables from the fact sheet pages would not be pulled into the database, nor would the information be ported over into the new site.

Also, in 2017, the STRSeq project was started [3]. STR sequence data from several large population studies [4-7] using commercially available kits was collected, processed, and submitted to GenBank. STRSeq GenBank records contain additional information geared toward the human identification community and are not normally found in GenBank records. The format and content of the modified records were the results of a collaborative effort between NIST and NCBI. The records are organized in nested BioProjects at NCBI to facilitate access. Flat files of the records can be downloaded individually, by locus, or by group: common autosomal, alternate autosomal, Y chromosome, X chromosome, or the entire set.

Present

In the last year, STRBase has undergone major publicly visible changes. The website has been reorganized and made more user-friendly. The updated STRBase continues to evolve by adding new projects (e.g., RGTM 10235, Forensic DNA Resource Samples -https://strbase.nist.gov/Information/RGTM_10235) and information (e.g., targeted STR locus sequences).

In addition, in the last year, the STRSeq records have also undergone their own format and content updates, incorporating the 2024 International Society for Forensic Genetics (ISFG) nomenclature recommendations [8]. These updates to the *human identification* section of the GenBank records include bracketing consistent with STRNaming [9], indication of the ISFG minimum reporting range, and incorporation of a sequence identification (SID) code [10]. As of the publication of this document, all the autosomal and X STR loci records have been updated to the new format of the STRSeq record, and Y STR record updates are ongoing.

The GenBank records can be downloaded as flat text files that are easily manipulated computationally but are not easily accessible for an overall review of multiple sequences at once. STRBase provides an ideal platform to include STR sequences in locus-specific sets. The new *Sequences* tab, which has been added to

the locus information (Figure 1), has been developed to include nomenclature and sequence information in an easy-to-search and view structure. Table 1 describes the information contained in the new *Sequences* tab on STRBase. The columns can be sorted or filtered. The data in these tables are being pulled from the database, which was expanded to incorporate this new data.

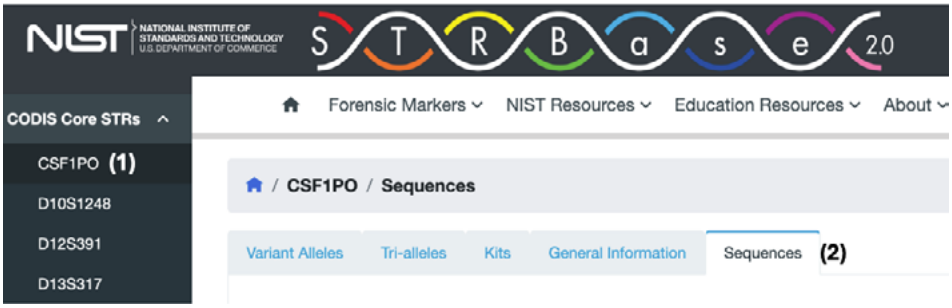


Figure 1. The new STRBase tab can be accessed through the locus information side menu (1). The newly added *Sequences* tab (2) location is also shown.

Table 1. Description of the content of the sequencing table in STRBase. The example is from the CSF1PO Sequences table.

Column Name	Content	Example
Accession	NCBI GenBank record accession	MK295183
Allele	Length-based allele call	6
Kit Codes	Kit names abbreviated; a legend is provided for conversion	FS[a]
STRNaming	Bracketed sequence	TCTA[6]ATCT[3]
Sequences	Sequence string	TAAGTACTTCCTATC-TATC…
SID Code	Sequence Identification code for the minimum range [8]	FEGJK
Coordinates	Reference genome GRCh38 coordinates	chr5:150076312..150076389

FS – ForenSeq DNA Signature Prep Kit

Future

The newly added sequence table has basic functionality, including sorting and filtering of records. Future work will incorporate alignment and highlighting of

sequences to better visualize the differences between the sequences in each table. Additional links to external information (i.e., GenBank records) will also be incorporated into the table. The database also holds additional information about these sequences, which could be provided in the table by request or default.

The identification of sequences will also be incorporated into the STRBase site using a STRSeq sequence search function. The goal is to allow users to submit a sequence string or FASTA file of sequences, along with additional information, such as the locus and the sequencing kit. The search function will return information about the sequences, such as the GenBank accession number (if one exists), sequence bracketing according to current recommendations, and the SID code for the minimum range, among other information. It will also identify sequences not contained in the STRSeq BioProject. This could open communication between the STRSeq project and the user to submit these new sequences to STRSeq, if appropriate.

The future plans for both STRBase and STRSeq are to maintain and expand, incorporating STRSeq information into STRBase to maximize usefulness for the human identification community. Feedback from the community is always welcome and can be sent to either strbase@nist.gov or strseq@nist.gov.

Disclaimer

Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. Certain commercial software, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

Research Protections Declaration

All work performed at NIST has been reviewed and approved by the National Institute of Standards and Technology Research Protections Office. This study was determined to be “not human subjects research” (often referred to as research not involving human subjects) as defined in U. S. Department of Commerce Regulations,

15 CFR 27, also known as the Common Rule (45 CFR 46, Subpart A), for the Protection of Human Subjects by the NIST Human Research Protections Office and therefore not subject to oversight by the NIST Institutional Review Board.

Conflict of interest

None.

References

[1] STRBase Archive, Accessed Oct. 7, 2024, <https://strbase-archive.nist.gov>

[2] STRBase, Accessed Oct. 7, 2024, <https://strbase.nist.gov>

[3] The STR Sequencing Project (human), BioProject Accession: PRJNA380127, Accessed Oct. 7, 2024, <https://www.ncbi.nlm.nih.gov/bioproject/380127>

[4] Novroski NMM *et al.,* «Characterization of genetic sequence variation of 58 STR loci in four major population groups.», *Forensic Sci Int Genet*, 2016 Nov; 25:214-226.

[5] Gettings KB *et al.,* «Sequence-based U.S. population data for 27 autosomal STR loci.», *Forensic Sci Int Genet*, 2018 Nov;37:106-115.

[6] Devesse L *et al.,* «Concordance of the ForenSeq system and characterization of sequence-specific autosomal STR alleles across two major population groups.», *Forensic Sci Int Genet*, 2018 May;34:57-61.

[7] Phillips C *et al.,* «Global patterns of STR sequence variation: Sequencing the CEPH human genome diversity panel for 58 forensic STRs using the Illumina ForenSeq DNA Signature Prep Kit.», *Electrophoresis*, 2018 Nov;39(21): 2708-2724.

[8] Gettings KB *et al*., "Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on short tandem repeat sequence nomenclature." *Forensic Sci Int Genet*, 2024, 68:102946.

[9] Hoogenboom J *et al.*, "STRNaming: generating simple, informative names for sequenced STR alleles in a standardized and automated manner." *Forensic Sci Int Genet*, 2021, 52:102473.

[10] Young B *et al.*, "A nomenclature for sequence-based forensic DNA analysis." *Forensic Sci Int Genet*, 2019, 42:14-20.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 443-446
DOI: https://dx.doi.org/10.15304/cc.2025.1869

The importance of staff genetic profiles in DNA databases

Vanessa Bogas[1*], Paula Cardoso[1], Ana Margarida Bento[1], Francisco Corte-Real[1,2], Pedro Brito[1]

[1] *Instituto Nacional de Medicina Legal e Ciências Forenses, Coimbra, Portugal.*
[2] *Faculdade de Medicina da Universidade de Coimbra, Portugal.*

** Corresponding author at: Polo das Ciências de Saúde (Polo III), Azinhaga de Santa Comba, 3000-548 Coimbra, vanessa.b.bogas@inmlcf.mj.pt, 00351 239 854 230*

Abstract
The Portuguese DNA profile database for civil and criminal investigation was approved in 2008 and by the end of 2023, had a total of 23775 genetic profiles. Around 30% of the profiles were from samples collected as evidence for criminal investigation purposes and 2% were staff profiles. The primary objective of DNA databases is to help criminal investigations, by detecting matches between profiles from samples collected at crime scenes and profiles from convicted offenders allowing the identification of the perpetrator who left biological traces at a crime scene.
Nearly 1% of the national matches in the Portuguese DNA database resulted between crimes scene samples and staff.
The aim of this work is to emphasize the importance of the inclusion of the staff profiles in the national DNA database, leading to identify possible cross-contaminations.

Keywords
Elimination DNA database; staff profiles.

Introduction

The Portuguese DNA Database Law was approved in 12th of February of 2008, for civil identification and criminal investigation purposes. The first DNA profile was imported in 2010 and by 31st of December of 2023 the DNA Database had a total of 23775 genetic profiles, 15900 of which were from convicted offenders, around 7212 profiles were from samples collected for criminal investigation and 564 from staff. The remaining 99 profiles are for civil identification purposes as well as volunteers.

The primary objective of DNA databases is to help criminal investigations, by detecting matches that may allow the identification of profiles from samples collected at crime scene as evidence, and to help civil identification through matches between profiles of unidentified human remains and profiles obtained from personal belongings of missing persons or relatives of missing persons.

Until the 31st of December of 2023, the Portuguese DNA database had produced 989 matches at national level, of which 60% were between profiles from crime scene evidence with convicted offenders, 39% were between profiles obtained from evidence samples of criminal investigation, and 1% were between profiles of collected evidence and staff profiles.

Methods

To carry out this study, all the matches occurred with staff profiles were considered since the Portuguese DNA Database went into production until 31st of December of 2023. Subsequently, using Excel software, the data were analyzed.

Results

Until 31st of December of 2023, the Portuguese DNA database counted with 564 staff profiles (**Figure 1a**). Since the first match with a staff profile in 2017, almost every year 1 match has occurred with staff profiles, leading to a total of 8 matches in the DNA database to this date **(Figure 1b)**, representing nearly 1% of the national matches. All of these matches occurred with profiles from samples collected in different crime scenes, three of these were single profiles, two of which with the same staff profile, and the other five were mixture profiles. One of the mixture profiles matched perfectly with the combination of two different staff profiles.

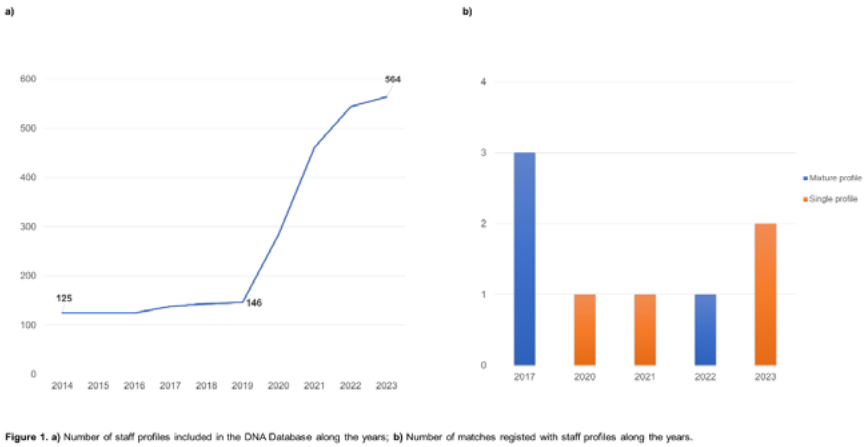


Figure 1. a) Number of staff profiles included in the DNA Database along the years; **b)** Number of matches registed with staff profiles along the years.

Discussion

Since the beginning of the Portuguese DNA database, 8 matches were detected between staff profiles and genetic profiles obtained from crime scene samples, which were assumed as evidence. However, because the profiles of staff that had contaminated these samples were already been entered into the DNA database, these evidence profiles were removed from the DNA database and no longer were considered for investigation purposes.

A higher number of staff profiles in the DNA database improves the detection of cross-contamination between staff profiles and profiles from samples collected in crime scenes.

Conclusion

To assure the integrity of an evidence and the reliability of the DNA profile obtained, every person who could introduce cross-contamination, whether they are responsible for collecting evidence at crime scenes or manipulating samples in the laboratory, must scrupulously comply with all contamination prevention guidelines. Contaminations can occur inadvertently, however, in accordance with ENFSI's

recommended guidelines, laboratories should have all their staff profiles inserted in a laboratory elimination DNA database and also in the national DNA database. Profiles of staff external to the laboratories who may in some way come into contact with the samples, should also be included in the national DNA database, enabling to detection of cross-contaminations which otherwise would not be detected [1].

Acknowledgments

None.

Conflict of interest statement

The author declare that they have no competing interests.

References

[1] ENFSI DNA Working Group, DNA Database Management. Review and Recommendations, April 2019

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 447-452
DOI: https://dx.doi.org/10.15304/cc.2025.1869

oYSTeR: A new Y-chromosomal ancestor estimator to improve speed, accuracy and versatility for forensic Y-DNA kinship investigations

Sofie Claerhout[1,2], Marc Wijnands[1,3], Jerry Hoogenboom[3], Heleen Coreelman[1,2], Titia Sijen[3,4], Ronny Decorte[2], Kristiaan J. van der Gaag[3,*]

[1] *CSY, Crime Scene Y-DNA Investigations, Interdisciplinary Research Facility IRF, KU Leuven KULAK, Kortrijk, Belgium.*

[2] *Forensic Biomedical Sciences, Department of Imaging & Pathology, KU Leuven, Leuven, Belgium.*

[3] *Division of Biological Traces, Netherlands Forensic Institute, The Hague, the Netherlands.*

[4] *Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, the Netherlands.*

** Corresponding author at: Division of Biological Traces, Netherlands Forensic Institute, Laan van Ypenburg 6, 2497 GB, The Hague, The Netherlands, k.van.der.gaag@nfi.nl*

Abstract

The Y-chromosome serves as a valuable indicator of paternal kinships in forensic DNA investigations. To reconstruct genealogies, the time to the most recent common ancestor (tMRCA) among paternal relatives can be estimated through in-depth analysis of Y-chromosomal short tandem repeats (Y-STRs). The previously developed 'YMrCA' calculator showed to be a remarkable advancement by including individual and haplogroup-specific Y-STR mutation rates, while accounting for multi-step changes and hidden, back, or parallel mutations. However, computational demands leading to lengthy runtimes (30 minutes per genealogical pair) and the absence of sequencing data integration, motivated us to enhance the existing calculator. A novel kinship estimating tool was developed, named 'oYSTeR' (Objective Y-Simulator To Estimate Relations). This advanced prediction model integrates the ability to include hundreds of Y-STRs, allele-specific mutation rates as well as Y-STR motif-specific mutation rates, and even Y-STR sequence data. Despite the integration of these additional Y-characteristics, oYSTeR is capable of estimating kinships for hundreds of generations without requiring a supercomputer. To do so, the calculator uses discrete convolutions, automatically accounting for parallel mutations. The oYSTeR prediction tool was validated using our extended CSY-database containing 1,120 biologically related males confirmed by 46 Y-STRs (capillary electrophoresis) covering 18,109 generations in total. Promising results showed that oYSTeR estimates 36,000 times faster than YMrCA with a runtime of 0.05

seconds per genealogical pair providing tMRCA estimations for all genealogical pairs. Currently, in terms of tMRCA estimation accuracy, we observed little difference with the YMrCA calculator. Given its dynamic, rapid, and user-friendly nature, oYSTeR stands as a future-proof calculator, ready to integrate additional Y-STR mutation rate data as they become available. With more accurate knowledge, for example on allele-specific mutation rates and complex Y-STR motif mutation rates, oYSTeR will be able to estimate kinships closer to reality. Therefore, oYSTeR offers the next step towards more objective tMRCA estimation for forensic DNA kinship investigations.

Keywords
tMRCA, MPS, Y-chromosome, Y-STR, Paternal kinships.

Introduction

The Y-chromosome is a useful marker for identifying paternal relationships in forensic DNA analysis. By examining Y-chromosomal short tandem repeats (Y-STRs), it is possible to estimate the time to the most recent common ancestor (tMRCA) among individuals related through the paternal line, aiding in the reconstruction of genealogies. *The previously developed 'YMrCA' calculator led to an advancement by including individual and haplogroup-specific Y-STR mutation rates, while accounting for multi-step changes and hidden, back, or parallel mutations. After validation, the YMrCA demonstrated a significantly improved success rate and reduced generation error compared to state-of-the-art kinship models [1]. However, computational demands leading to lengthy runtimes (30 minutes per genealogical pair) and the absence of sequencing data integration, motivated us to enhance the existing calculator to be able to process data of large Y-STR (sequence) panels such as the CSYseq [2].*

A collaboration between CSY-Belgium (KU Leuven) and the Netherlands Forensic Institute (NFI), has resulted in the development of a novel kinship estimating tool, named 'oYSTeR' (Objective Y-Simulator To Estimate Relations). This advanced prediction model seamlessly integrates the ability to include hundreds of Y-STRs, allele-specific mutation rates as well as Y-STR motif-specific mutation rates, and even Y-STR sequence data by using a discrete convolution rather than a limited (due to computational limits) number of simulations.

In this study, oYSTeR performance was tested on available data from 1,120 biologically related males confirmed by 46 Y-STRs covering 18,109 generations in total and compared to existing YMrCA predictions [3].

Material studied, methods, techniques

Implemented mathematical principles

oYSTeR was developed in Python (with an alternative faster version programmed in Julia). To decrease computational demand and increase analysis speed, a discrete convolution [4] was implemented instead of a fixed number of simulations as used in the YMrCA. The basic principle of the discrete convolution is displayed in **figure 1**. For readability, the visualization in **figure 1** is limited to only a few options, including -1, 0 and +1 as starting repeat length of alleles at a given locus and -1, 'X' (no mutation) and +1 for mutation rates. However, the mutation options in the incorporated convolution can be extended for any mutation (e.g. multistep mutations, back or parallel mutations) while encountering multiple mutation characteristics (e.g. CE length based mutation rates or rates defined per sequence repeat block).

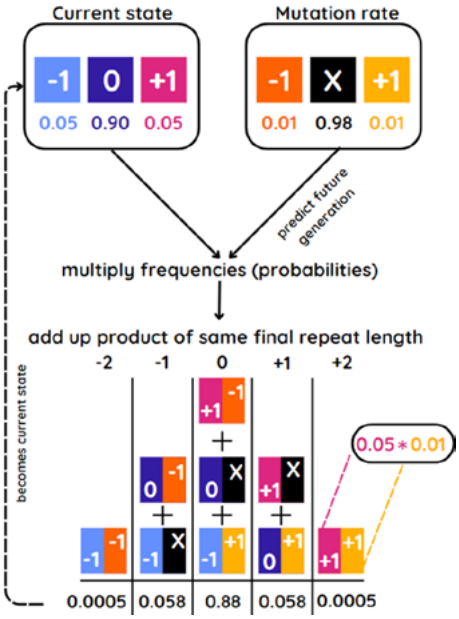


Figure 1. *Illustration of the basic principle behind the implemented discrete convolution. For each allele length of each meiosis distance, the probability of occurrence is calculated based on the starting variation of the previous meiosis (left top) and the mutation rates (right top). The probabilities of each separate length are summed up to determine the final probabilities of each length after the next meiosis and is repeated for each meiosis step.*

Credibility interval

After predicting the probabilities for each meiosis distance (absolute and relative, with relative meaning all calculated meiosis probabilities scaled to a total of one), 95%, 75% and 50% credibility intervals (CI) are determined. These define the range of meioses with the highest relative probabilities summing up to either 95%, 75% and 50% (see **figure 2** for a prediction curve example).

Input and output formats

For the genotypes, a tab-separated input file is required containing a header row with Y-STR marker names and two lines with the Y-haplotypes of the individuals for which the tMRCA needs to be calculated (multiple pairs can be included by separating each pair by an empty row). The contents of the library file (mutation rates) and resulting output file are displayed in table 1 (each line in the table is a column of the tab separated file).

Table 1. Contents of library and output file for oYSTeR analysis.

Library file	Sample file (with header of Y-STR names)
Marker name, for MPS with the addition '[repeat unit]_[repeat occurrence]'	Follow up number of genotype pair
CE/MPS (data type category per marker)	95% CI
Mutation interval boundary (for example '10')	75% CI
Minimum and maximum length (for example '1;100' if the minimum length of the unit is 1 repeat and the maximum length is 100 repeats)	50% CI
Mutation rate before interval boundary (for example '0.0001' meaning 1 mutation in every 10,000 meioses when the allele length was ≤10 repeats)	Mode
Mutation rate after interval boundary (for example '0.0002' meaning 1 mutation in every 5,000 meioses when the allele length was >10 repeats)	Relative probabilities for each meiosis distance (displayed as '[prob meiosis 1, prob meiosis 2, prob meisosis 3, …]'
	Absolute probabilities for each meiosis distance (displayed as '[prob meiosis 1, prob meiosis 2, prob meiosis 3, …]

Output visualizations

In addition to the predictions, several visualizations are created from the probabilities as displayed in figure 2. The visualization file (html-format) also displays separate probabilities for a meiosis distance of ≤1, ≤2, ≤3, ≤4, ≤5 and ≤10 that can indicate whether close relatedness can be (almost) excluded or not (not displayed in figure).

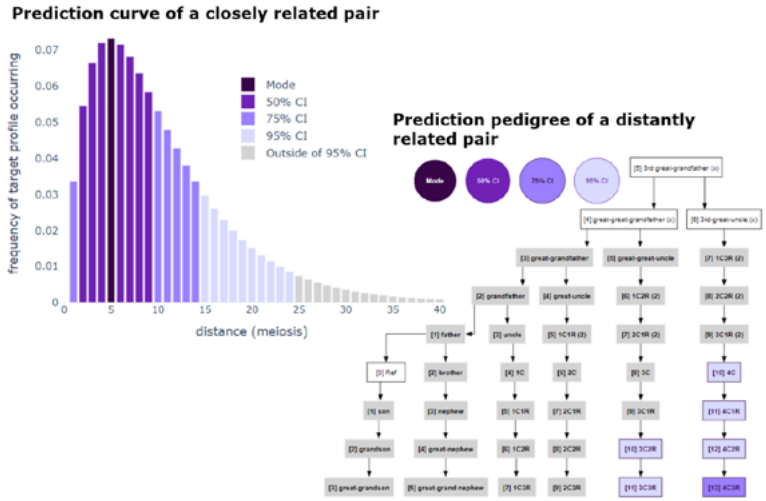


Figure 2. *oYSTeR visualizations that are created for each sample pair. The prediction curve is displayed for a closely related pair and the prediction pedigree for a more distantly related pair.*

Results / Discussion

The oYSTeR prediction tool was verified using our enlarged CSY-database, comprising 1,120 biologically related males identified through 46 Y-STR markers (via capillary electrophoresis), encompassing a total of 18,109 generations. The encouraging findings revealed that oYSTeR computes 36,000 times quicker than YMrCA [1], with a processing duration of only 0.05 seconds per genealogical pair. It successfully generated tMRCA predictions for every genealogical pair, showcasing its capability to accurately determine relationships between males, even in cases involving numerous Y-mutations or larger multi-step variations. Regarding tMRCA estimation precision, we noticed minimal discrepancies when compared to the YMrCA calculator. Nevertheless, further comprehensive investigation of Y-STR mutation rate dynamics and the inclusion of genealogical pair Y-STR sequence data remains necessary to assess its impact on tMRCA prediction accuracy.

Visualizations of the oYSTeR predictions serve multiple purposes. First, the shape of the prediction curve helps to understand that for a broad curve, chances can be similar for multiple meioses distances (as is often the case) to focus more on CI ranges rather than just the mode of the curve. The pedigree displays the most obvious candidates from a family tree (relative to the first profile) for which the probability of the compared profile is the most obvious. For predictions that point

to a more distant relatedness, it can be informative to use the table with probabilities for close relatedness to indicate if these can actually be (almost) excluded or not or if there is still a small chance for close relatedness as well.

It should be noted that the probabilities are relative to the number of meioses that were included in the calculation. Especially for more distant relatedness, probabilities and CI ranges will only be reliable if a sufficient number of meioses are included in the calculation (this can be configured to preference) as can easily be recognized from the outer probabilities of the prediction curve.

Conclusion

oYSTeR is a future proof Y-kinship estimation tool given its dynamic, rapid, and user-friendly nature, *enabling fast and accurate analysis of larger Y-STR datasets suited for both CE as well as sequence data.*

Acknowledgments

Special thanks to Klaas Slooten for assisting in the mathematics of the prediction model. Funding for this research was provided by FWO (SC; 1265622N).

Conflict of interest statement

The authors disclose no conflict of interest.

References

[1] Claerhout S, Vanpaemel S, Gill MS et al. YMrCA: Improving Y-chromosomal ancestor time estimation for DNA kinship research. Hum Mutat. 2021 Oct;42(10):1307-1320.

[2] Claerhout S, Verstraete P, Warnez L et al. CSYseq: The first Y-chromosome sequencing tool typing a large number of Y-SNPs and Y-STRs to unravel worldwide human population genetics. PLoS Genet. 2021 Sep 7;17(9):e1009758.

[3] Claerhout S, Vandenbosch M, Nivelle K et al. Determining Y-STR mutation rates in deep-routing genealogies: Identification of haplogroup differences. Forensic Sci Int Genet. 2018 May;34:1-10.

[4] Heckbert PS. Fourier Transforms and the Fast Fourier Transform (FFT) Algorithm. Mathematics, Physics 1998.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 453-459
DOI: https://dx.doi.org/10.15304/cc.2025.1869

A comparison of likelihood ratios between STR sequencing and capillary electrophoresis for complex mixtures

Maja Sidstedt[1*], Arvid H. Gynnå[1], Johannes Hedman[1,2], Ronny Hedell[1]

[1] *National Forensic Centre, Swedish Police Authority, Linköping, Sweden*

[2] *Applied Microbiology, Department of Chemistry, Lund University, Lund, Sweden*

** Corresponding author at: Brigadgatan 13, 581 94 Linköping, maja.sidstedt@polisen.se*

Abstract

Wide application of massively parallel sequencing (MPS) in forensic DNA profiling has been hampered by limited knowledge on the possible information gain compared to the established capillary electrophoresis (CE) methods. MPS analysis of short tandem repeats (STRs) offers increased discriminatory power due to the inclusion of sequence variants in addition to the often higher number of STR markers. In this study, the Mainstay SE sequencing kit (28 autosomal STRs, Qiagen) was compared to the CE assay PowerPlex Fusion 6C (23 autosomal STRs, Promega), with respect to likelihood ratios (LR) following analysis of mixtures. Specifically, a set of 19 complex mixtures with two to four contributors in various proportions (1:9, 1:1:8, 1:1:18, 1:1:2:8 and 1:1:2:18) were analyzed with both methods. LR calculations were performed using the continuous model EuroForMix for both sequencing and CE results while applying various filters. This study contributes with knowledge regarding when sequencing is preferable over CE analysis, thus aiding forensic laboratories in the decision to implement STR sequencing.

Keywords

MPS, Mixtures, Probabilistic genotyping, STR, Likelihood ratio.

Introduction

Massively parallel sequencing (MPS) of short tandem repeats (STRs) offers increased discriminatory power due to sequence variants. Together with the possibility of analyzing higher numbers of STR markers, this may generate increased weights of evidence for true contributors compared to the established capillary

electrophoresis (CE) methods [1]. However, the wide application of MPS STR analysis has been hampered by the limited knowledge on the possible information gain compared to CE. In this study, we compared the likelihood ratios for a set of complex mixtures obtained with either MPS or CE with the aim to contribute with knowledge regarding when sequencing is preferable over CE.

Methodology

The Mainstay SE sequencing kit with the ePCR1 buffer (28 autosomal STRs, Qiagen) was compared to the CE assay PowerPlex Fusion 6C (23 autosomal STRs, Promega), with respect to likelihood ratios (LR) following analysis of mixtures. Specifically, a set of 19 complex mixtures with two, three or four contributors in various proportions (1:9, 1:1:8, 1:1:18, 1:1:2:8 and 1:1:2:18) and the minor contributor DNA amount at 32.5-167 pg was analyzed with both assays in triplicate.

LR calculations were performed using the continuous model EuroForMix (EFM) v. 4.0.8 for both CE and MPS results, applying various filters and settings (one for CE and two for MPS) (Table 1). For CE, in-house routine analysis settings and thresholds were used. Additionally, a provisional analysis using MPSProto v. 0.9.5 was performed for MPS results [2]. The hypotheses used were H1: DNA is from Person of Interest + K unknowns and H2: DNA is from K+1 unknowns. In total, 240 true contributor LR calculations per method were performed for the 19 mixtures, considering two or three replicates and with or without conditioning on the major.

Table 1. Overview of the applied STR typing assays, analysis settings and thresholds.

Assay	Fusion	Mainstay SE		
	CE - EFM	Stringent - EFM	Minus2 - EFM	MPSproto
Analysis instrument	ABI3500xL	MiSeq FGx	MiSeq FGx	MiSeq FGx
Software	GeneMapper	UAS[a]	UAS[a]	UAS[a]
Analytical threshold	85-140 RFUs[b]	1.5% of total reads per locus and ≥ 30 reads	≥ 30 reads	≥ 11 reads

Stutter filter	Yes[c]	Yes[d]	Only n-2 stutter	No
Stutter model (EFM/MPSproto)	No	No	Yes	Yes
No. of STR markers	23	28	28	27 (excluding SE33)
Allele designation	Nr. of repeats	LUS+ [3]	LUS+ [3]	Forward bracket [2]
Allele frequency	NIST 1036 Caucasian (n=361) with lowest allele frequency of 0.5%			

a Universal Analysis Software (Qiagen)
b Relative Fluorescence Units
c according to manufacturer recommendations, but using four standard deviations instead of three
d according to manufacturer recommendations for ePCR buffer

Results and Discussion

We expected that the sequencing results would yield higher LRs than CE for the complex mixtures given the advantage of a higher number of markers and more possible alleles. However, when using low analytical thresholds (below 30 reads) together with the EFM stutter model in initial test, MPS analysis gave no usable results (data not shown). This was due to that the model does not handle n-2 stutters or other minor artefacts. Instead, more stringent thresholds were required. Thus, the methods called "Stringent" and "Minus2" were applied (Table 1). Unfortunately, these interpretation methods resulted in impaired limits of detection for minor contributor alleles (Figure 1). Most of the lost information was due to the high stutter filter and the proportional analytical threshold of 1.5% in the "Stringent" method, as the "Minus2" method resulted in lower proportions of allele dropouts for all numbers of contributors (NOCs).

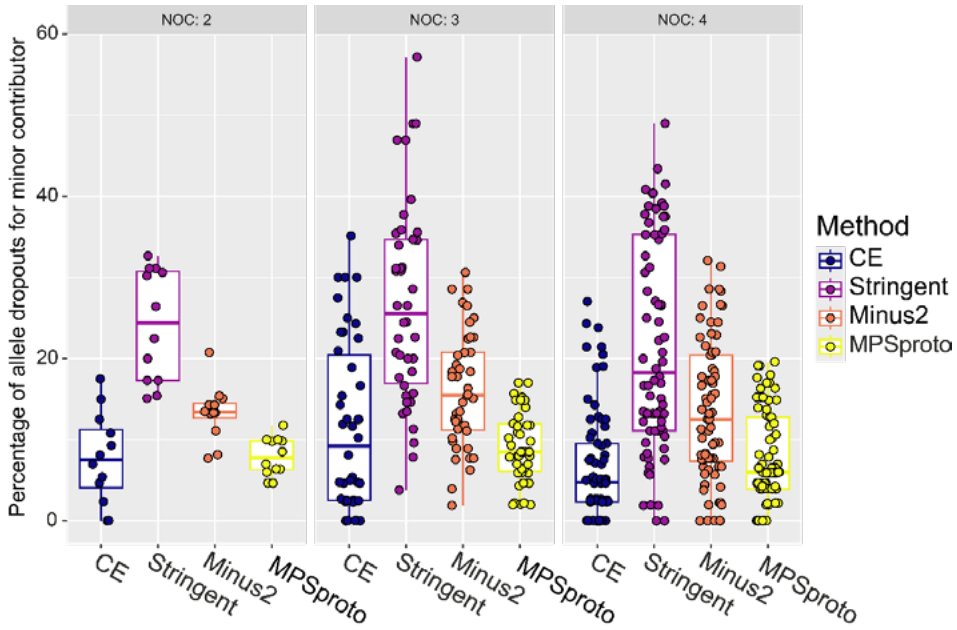


Figure 1. Proportion of allele drop-outs for minor contributors in DNA mixtures when using filters applied as described in Table 1. Each datapoint represents the proportion of missing alleles for respective minor contributor in each mixture, grouped into either two-, three- or four contributors (NOC 2-4).

Nevertheless, the LR values were similar for CE and sequencing for two and three contributor mixtures (Figure 2). For NOC:2 and the "Minus2" method, although more alleles were detected, there were some calculations that yielded lower LRs compared to the other methods. This was most likely due to compound stutter artefacts that could not be explained by the EFM stutter model. With the more complex four contributor mixtures, there was a slight advantage for the sequencing results since CE resulted in more false negative LR values (Figure 2). Note that these CE results would have been manually reported as inconclusive in casework. Despite having stringent filter settings for MPS results, we thus observed similar LRs for both CE and MPS, which is in line with the findings of Benschop et. al [4].

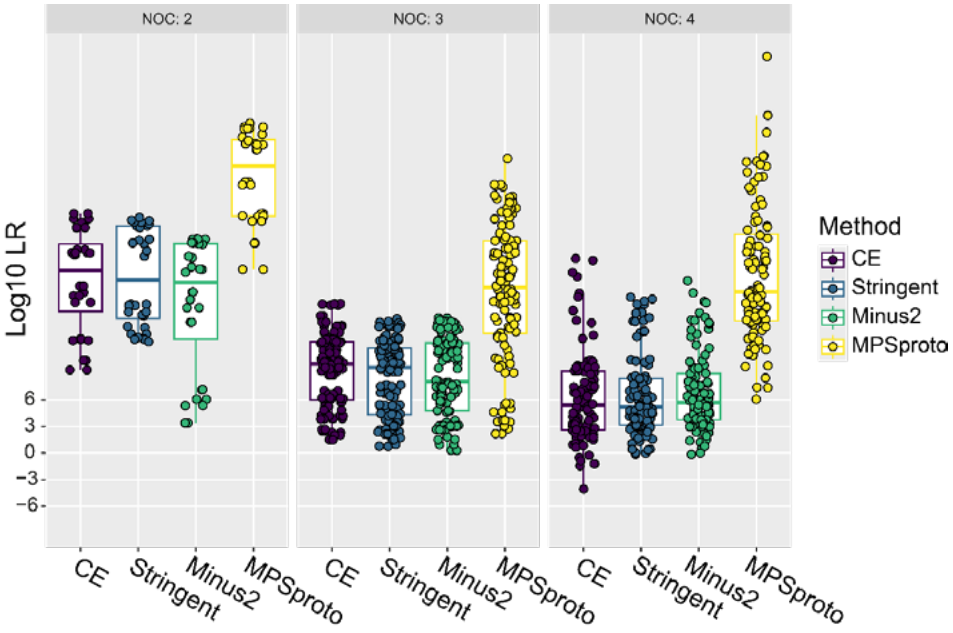


Figure 2. Likelihood ratios (log10) grouped according to the NOC for each method with analysis settings as presented in Table 1. Values above 0 support H1 (true hypothesis) and values below 0 support H2 (false hypothesis). Calculations were performed with either two or three replicates and with or without conditioning on the major contributor.

Since a quite stringent filter was necessary for the sequencing data, we hypothesized that an improvement could be made by using models developed specifically for sequencing, such as MPSproto [2]. Here, we used a limited dataset for calibration and therefore this should be viewed as a provisional evaluation. Very promising results were achieved with MPSproto, yielding LRs above 10^6 for even the most challenging NOC:4 mixtures (Figure 2 and Figure 3).

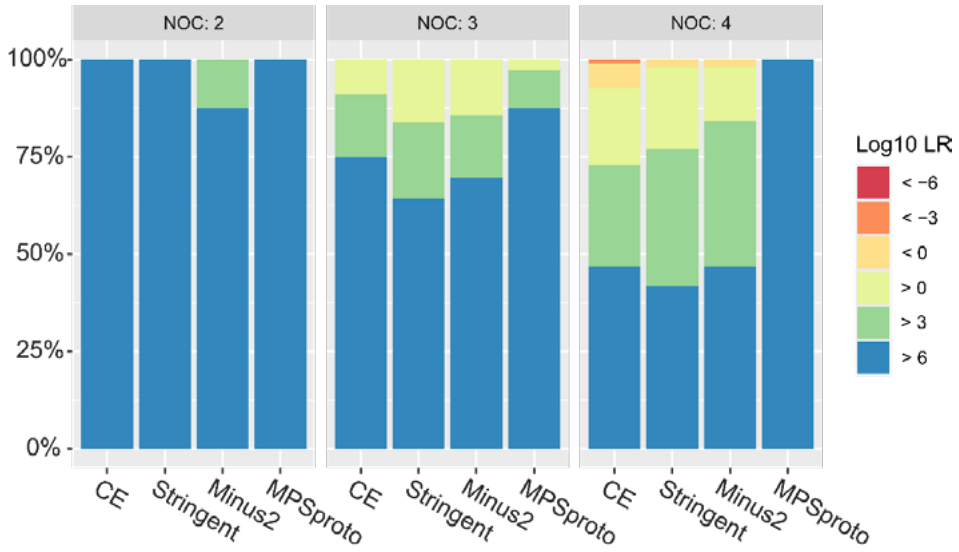


Figure 3. Proportion of Log10 LR values that fall into the intervals delimited by -6, -3, 0, 3, 6 for either two, three or four contributors. Values of Log10 LR above 0 support H1 (true hypothesis) and values of Log10 LR below 0 support H2 (false hypothesis).

Conclusion

Our results suggest that for STR sequencing to add value in weight of evidence calculations for mixtures there is a need for advanced modeling of sequencing specific artefacts and noise. Very promising results were obtained with MPSproto.

Acknowledgments

The authors are grateful to Monica Giuffrida (University of Copenhagen) for sharing analysis settings for the SE33 locus.

Conflict of interest statement

None

References

[1] Ballard, D., Winkler-Galicki, J., Wesoły, J. (2020). Massive parallel sequencing in forensics: advantages, issues, technicalities, and prospects. Int. J. Legal Med., 134, 1291-1303.

[2] Bleka, Ø., Just, R., Agudo, M.M., Gill, P. (2022). MPSproto: An extension of EuroForMix to evaluate MPS-STR mixtures. Forensic Sci. Int. Genet., 61, 102781.

[3] Just, R.S., Le, J., Irwin, J.A. (2020). LUS+: Extension of the LUS designator concept to differentiate most sequence alleles for 27 STR loci. Forensic Sci. Int.: Reports, 2, 100059.

[4] Benschop, C.C.G., van der Gaag, K.J., de Vreede, J., Backx, A.J., de Leeuw, R.H., Zuñiga, S., Hoogenboom, J., de Knijff, P., Sijen, T. (2021). Application of a probabilistic genotyping software to MPS mixture STR data is supported by similar trends in LRs compared with CE data. Forensic Sci. Int. Genet., 52, 102489.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 461-465
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Evaluation of uncertainties in investigations of different type of kinship scenarios

Eszter Rockenbauer[1*], Carmen Tomas[1], Charlotte Hallenberg[1]

[1] *Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark.*

** Corresponding author: eszter.rockenbauer@sund.ku.dk, tel: +4535326284*

Abstract
In this study we investigated the resolution power of the *GlobalFiler™ PCR Amplification kit in 31 kinship scenarios.* We performed *50.000 simulations for each scenario and visualized the results by plotting the log of the* likelihood ratio *values for the true hypotheses, the null hypothesis (H0) and the alternative hypothesis (H1). The scenarios were categorized as either conclusive, intermediate or inconclusive. Improved resolution was observed when additional relatives were introduced to the tested scenarios. This work underlines the importance of simulation studies when handling deficiency cases.*

Keywords
Likelihood ratio, simulation studies, *deficiency* cases

Introduction

Most currently used autosomal STR kits are well-suited for simple kinship investigations (i.e.: parent/child duos or trios), where high likelihood ratio (LR) values (above 100.000 to 1) are usually achieved. For deficiency cases however, LR values obtained with autosomal STR kits alone are often in a much lower range. In this study, we aimed to investigate within which LR ranges the results ought to be considered as inconclusive for the kinship scenarios most often requested to our laboratory.

Techniques

Simulated pedigrees for 31 kinship scenarios, listed in Table 1, were generated using Familias software v. 3.3.1 (www.familias.no) [1-2]. The simulations were based on an allele frequency database for the Danish population for the 21 autosomal STRs included in the GlobalFiler™ kit (Thermo Fisher Scientific). For each kinship scenario 50.000 simulations were performed. The logarithm of the LR values for each scenario were plotted in two density curves, for true H0 and for true H1, respectively. The inconclusive area was defined as the region with more than 0.1 % true H0 and true H1 with overlapping LRs, from now on referred to as "the overlapping LR range".

Table 1. Overview and classification of the 31 kinship scenarios with simulated pedigrees. Following values are shown, percentage of LR values of each true hypothesis within the overlapping region, median LR values for true H0 and true H1. Color coding: Green for conclusive, Yellow for intermediate, Red for inconclusive. The scenarios are listed in order of most to least informative.

No.	H0	H1	% H0 and H1 within the overlap	Median H0	Median H1
1	Mother-child-man	Mother-child	0	4,54E+11	2,27E-33
2	Mother-child-childs 3 full sibs	Mother-child-childs 3 half sibs	0	1,67E+09	2,04E-18
3	Child-man	Unrelated	0	1,16E+08	1,54E-24
4	Mother-child-man's parents	Mother-child	0	3,82E+07	2,03E-20
5	Mother-child-childs 2 full sibs	Mother-child-childs 2 half sibs	0	3,11E+07	4,10E-12
6	Child-childs 3 full sibs	Child-childs 3 half sibs	0	5,75E+06	3,61E-13
7	Mother-child-man's 3 full sibs	Mother-child	0,626	8,86E+05	6,48E-10
8	Mother-child-childs 3 half sibs	Mother-child	0,651	1,82E+07	2,30E-09
9	Child-man's parents	Unrelated	0,652	7,07E+04	2,67E-12
10	Child-childs full sib	Unrelated	3,346	6,26E+06	9,30E-06
11	Child-childs 2 full sibs	Child-childs 2 half sibs	6,016	4,54E+04	1,03E-07
12	Mother-child-childs 2 half sibs	Mother-child	7,198	7,36E+04	2,72E-06
13	Mother-child-man's 2 full sibs	Mother-child	7,512	6,17E+04	3,74E-06
14	Child-childs 3 half sibs	Unrelated	10,477	7,32E+03	1,37E-06
15	Child-man's 3 full sibs	Unrelated	11,246	6,58E+03	2,57E-06
16	Child-childs 2 half sibs	Unrelated	23,21	1,33E+03	2,80E-04
17	Child-man's 2 full sibs	Unrelated	23,556	1,26E+03	3,37E-04
18	Mother-child-childs full sib	Mother-child-childs half sib	30,225	2,21E+04	7,16E-04
19	Mother-child-man's single parent	Mother-child	51,122	8,07E+02	4,70E-03
20	Mother-child-childs half sib	Mother-child	51,219	7,98E+02	4,82E-03
21	Mother-child-man's full sib	Mother-child	51,5	7,70E+02	4,94E-03
22	Child-childs full sib	Child-childs half sib	58,57	9,69E+01	2,58E-02
23	Child-man's single parent	Unrelated	71,487	8,34E+01	2,30E-02
24	Child-childs half sib	Unrelated	71,599	8,00E+01	2,37E-02
25	Mother-child-man's 2 half sibs	Mother-child	79,894	1,73E+01	8,66E-02
26	Mother-child-man's 3 half sibs	Mother-child	81,739	3,09E+01	5,32E-02
27	Child-man's full sib	Unrelated	83,624	7,66E+01	2,46E-02
28	Child-man's 3 half sibs	Unrelated	88,182	8,22E+00	1,44E-01
29	Mother-child-man's half sib	Mother-child	92,137	5,89E+00	2,03E-01
30	Child-man's 2 half sibs	Unrelated	92,168	5,79E+00	1,98E-01
31	Child-man's half sib	Unrelated	96,733	2,96E+00	3,44E-01

Results and Discussion

The 31 scenarios (Table 1) were classified as conclusive, intermediate or inconclusive based on the following three parameters: the percentage of true H0 and true H1 within the overlapping LR range, the median LR value of true H0 and the median LR value of true H1 (see Table 2).

Eight scenarios (numbered 1 to 8 in Table 1) fell in the conclusive category. The percentage of true H0 and true H1 in the overlapping LR range for all scenarios was below 1% (i.e.: the risk of making a false conclusion based on the LR values obtained from 21 autosomal STRs alone is less than 1%). The median LR values for the true H0 hypotheses were all above 100.000 to 1 and for the true H1 hypotheses were below 1 to 100.000.

Ten of the 31 scenarios (numbered 9 to 18 in Table 1) fell in the intermediate category. For eight of them, the percentage of true H0 and true H1 in the overlapping LR range varied between 1% and 25%. The median LR values for the true H0 hypotheses for scenarios in this category ranged between 100.000 to 1 and 1.000 to 1 (in 9 out of the 10 scenarios). In all cases, true H1 hypotheses were below 1 to 10.000. The three scenarios (9, 10 and 18 in Table 1) were included in the intermediate category, despite borderline values.

The remaining 13 scenarios (19 to 31 in Table 1) fell in the inconclusive category with a percentage of true H0 and true H1 in the overlapping LR range above 50%. The median LR values for the true H0 hypotheses for all scenarios in the inconclusive category were below 1.000 to 1 and for true H1 above 1 to 1.000.

Table 2. On overview of the parameters used to classify the 31 kinship scenarios with simulated pedigrees listed in Table 1. Color coding: Green for conclusive, Yellow for intermediate, Red for inconclusive.

Classification	% H0 and H1 within overlapping LR range	Median H0	Median H1
Conclusive	< 1	> 1,0E+5	< 1,0E-5
Intermediate	1 to 25	1,0E+5 to 1,0E+3	1,0E-5 to 1,0E-3
Inconclusive	> 25	< 1,0E+3	>1,0E-3

We observed improvements of resolution, when one or more additional relatives were introduced to the tested scenarios. Three examples are illustrated in Figure 1, in top, middle and bottom row, respectively.:

1. When the child's biological mother is added to the scenario where the alleged paternal grandparents are tested (scenarios numbered 9 and 4 in Table 1)
2. When additional siblings of the child are included (scenarios numbered 22, 11 and 6 in Table 1)
3. When additional full siblings of the alleged father are introduced (scenarios numbered 21, 13 and 7 in Table 1).

As shown in Figure 1, adding a relevant individual to the scenario reduces the percentage of true H0 and true H1 in the overlapping LR range (illustrated with the area "1"). Also, an increase in the difference between the median values of the true hypotheses, and in the percentage of true H0 with LR values higher than 100.000 to 1 was observed. In summary, testing additional relatives, reduces the risk of making a false call and increases the expected LR values.

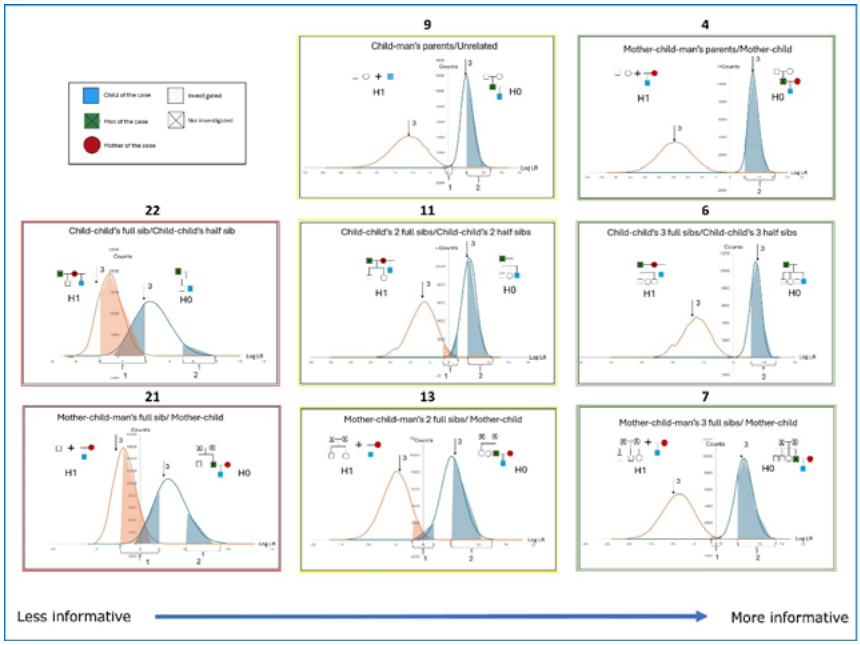


Figure 1: Density curves for log LR values (true H0 on the right, and true H1 on the left) calculated for simulated scenarios (numbered according to Tabel 1). The following is shown: The proportion of LR values of each true hypothesis within the overlapping area (1), proportion of LR values >100.000 to 1 (2), median LR values for true H0 and true H1 (3), and the pedigrees for true H0 and true H1 (see top left of the figure for more detailed explanation).

Conclusion

The 8 scenarios in the conclusive category can be recommended for kinship testing in our laboratory, using our standard autosomal STR kit alone. The 10 scenarios in the intermediate category can also be recommended, although, depending on the constellation, supplementary markers (Y-STRs, X-STRs, autosomal SNPs) might be needed. The remaining 13 scenarios, which are in the inconclusive range, need to be carefully evaluated. For some of these scenarios, our recently implemented MPS kit for microhaplotypes [3] could reduce the expected inconclusive range. Another way to reduce the inconclusive range is to include additional relevant relatives, if available. With this knowledge, we have a better picture of what to expect in a certain scenario and we can better guide and inform our costumers beforehand.

Conflict of interest statement

None.

References

[1] D. Kling, A. O. Tillmar, T. Egeland, Familias 3-Extensions and new functionality, Forensic Science International Genetics 13 (2014) 121-127.

[2] T. Egeland, P. Mostad, B. Mevåg, et al., Beyond traditional paternity and identification cases. Selecting the most probable pedigree. Forensic Science International 110 (2000) 1.

[3] C. Tomas, P. Rodrigues, C.G. Jønck, Z. Barekzay, et al., Performance of a 74-Microhaplotype Assay in Kinship Analyses, Genes 15 (2024) 224.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 467-473
DOI: https://dx.doi.org/10.15304/cc.2025.1869

A general approach for combining non-genetic and genetic data to solve complex kinship cases

Franco Marsico[1] and Thore Egeland[2,3]

[1] *Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina.*
[2] *Norwegian University of Life Sciences, Norway.*
[3] *Oslo University Hospital, Norway*

** Corresponding author: franco.lmarsico@gmail.com, +541149385494*

Abstract
The missing persons search is a complex process that requires the integration of both genetic and non-genetic information. The search involves comparing data from two entities, the Missing Persons (MPs), that have known identities but unknown whereabouts, and the Unidentified Persons (UPs) that may be individuals with unknown identities or unidentified remains. We introduce a mathematical framework that combines genetic data, such as short tandem repeats (STRs), with non-genetic traits like pigmentation characteristics (hair, skin, and eye color). This approach utilizes Likelihood Ratio (LR) models to evaluate different hypotheses (whether the UP is the MP or unrelated). By integrating both types of data, the method enhances the search process, thereby reducing the chances of false positives and negatives in LR based classification during database searches. Despite current challenges, such as the need for accurate population databases and potential correlations between genetic and non-genetic data, this integrated model offers a more reliable solution than other approaches, such as filtering, paving the way for further refinement and broader application in real-world applications.

Keywords
Forensic identification, Likelihood ratio, Pigmentation traits, Missing person identification, Genetic and non-genetic evidence.

1. Introduction

Forensic investigations into missing persons cases present complex challenges that demand the integration of both genetic and non-genetic data [1]. A Missing Person

(MP) is an individual with a known identity but unknown whereabouts, while an Unidentified Person (UP) may be a living individual with unknown identity or unidentified human remains. The search for MPs involves gathering testimonies, legal documents, family interviews, social media information, and medical records. For UPs, the focus is on analyzing remains and their context of discovery. Recent advancements in DNA analysis have enhanced forensic capabilities by allowing the inference of phenotypic characteristics such as skin, eye, and hair color, as well as age, among others [2].

Interpreting forensic evidence relies heavily on statistical methods such as the Likelihood Ratio (LR) approach, which quantify the weight of the evidence under different hypotheses (typically whether the UP is the MP or is unrelated). DNA evidence plays a crucial role, often involving the analysis of samples from UPs and genetic material from relatives of MPs, requiring accurate knowledge of family relationships represented by pedigrees. Complex scenarios like Disaster Victim Identifications (DVI) and large-scale Missing Person Identifications (MPI) present additional challenges, such as low statistical power due to scarce genetic data and the need to analyze multiple pedigrees simultaneously, leading to potential false positives [3]. To address these issues, we applied a recently developed mathematical framework [4, 5] that evaluates the statistical weight of different non-genetic sources of evidence. For example, it allows analyzing pigmentation trait evidence using a likelihood-ratio approach that accounts for conditional dependencies among traits. Recognizing that characteristics like eye color, hair color, and skin color are interrelated, the model presents a step forward compared to previous methods that assumed independence among these variables. Here, we show the impact of adding these models into the search in cases where few or only distant relatives of the missing person are available for genotyping. Furthermore, we discuss unsolved challenges for future developments.

2. Methods

2.1 Data

The presented methodology integrates both genetic and non-genetic data collected during the search process in missing person cases. Non-genetic data include physical traits such as biological sex, age, and pigmentation characteristics (hair color, skin color, and eye color). These data are obtained from various sources like DNA phenotyping, legal documents, testimonies from relatives and witnesses,

social media, and direct observation. These traits are represented as variables for sex (*S*), age (*A*), and a composite variable *C* for pigmentation traits. Genetic data are collected from UPs and relatives of MPs, involving genetic profiles analyzed through 23 STRs markers. Both types of data are organized into databases for UPs and MPs, allowing comparisons and analyses.

2.2 Likelihood Ratio models

To statistically evaluate the weight of the non-genetic evidence, we employed LR models previously developed [4, 5] that quantify the probability of observing the data under two competing hypotheses: the UP is the MP (H_1), and the UP is unrelated to the MP (H_2). The general formulation of the LR is the ratio of these probabilities. For multiple pigmentation traits, we used an extended model to account for the conditional dependencies among hair color, skin color, and eye color [5]. By using the chain rule of conditional probabilities, we calculate the joint probability of these interrelated traits, leading to a more accurate LR that reflects their combined effect on the identification process. All these models are implemented in the mispitools package [6], freely available on CRAN. For genetic evidence, we used the forrel package [7], that facilitates computation of LRs in missing person cases.

2.3 Computational simulations and performance evaluation

We performed computational simulations to assess the performance of our LR models, using mispitools [6]. These simulations involve generating synthetic datasets under both H_1 and H_2 scenarios to estimate error rates and evaluate metrics such as true-positive and false-positive rates. By simulating a large number of cases, we were able to evaluate the performance of the search, analyzing the LRs distribution under the different scenarios. We computed the Matthews Correlation Coefficient (MCC), for an LR threshold of 1 in order to evaluate the performance.

3. Results

In this section, we present the application of the LR models that combine genetic and non-genetic data to improve the search for missing persons.

3.1 Statistical power of DNA-based identification

We first analyzed a case where only one relative, a first cousin, was available for genotyping (Figure 1A). Using 23 STRs and conducting 10,000 simulations, we

observed overlapping Log10(LR) distributions for H_1 (UP is the MP) and H_2 (UP is not the MP). This overlap indicates that similar Log10(LR) values were obtained whether the UP was actually the MP or not when compared to the pedigree. Based on these distributions, MCC = 0.62, indicating low discrimination capability.

3.2 Addressing Pigmentation traits

Figure 1B illustrates the probability distributions of the combined pigmentation traits under H_1 for each characteristic. We considered a case where MP pigmentation traits are: Skin tone 3, with brown hair (color code 2) and brown eye color (code 2).

Higher probabilities are obtained when the pigmentation traits between the UP and MP coincide. However, errors such as misclassifications, data entry mistakes, and uncertainties in testimonies can lead to a non-zero probability to observe different traits between UP and MP even when H_1 is true. For LR computations, probability distribution considering H_2 as true, is obtained from the reference population. In this case, a synthetic simulated population was considered as a toy example. Having the probability distributions for H_1 and H_2 allowed computing the LR for each combination of traits (Supplementary Figure 1).

3.3 Combining genetic and non-genetic data

We applied our combined LR approach to analyze both genetic and non-genetic data. In cases where only distant relatives were available (e.g., second cousins), the genetic evidence alone was insufficient, resulting in overlapping LR distributions under both hypotheses. When pigmentation traits were included, the LR distributions became more distinct, allowing for better differentiation between matching and non-matching individuals (Figure 1C).

By integrating genetic and non-genetic evidence, we observed a clearer separation between the LR distributions under the two hypotheses. This integration reduces the risk of false positives (high LR values when H_2 is true) and false negatives (low LR values when H_1 is true), thereby enhancing the performance of the identification process, with a MCC = 0.91.

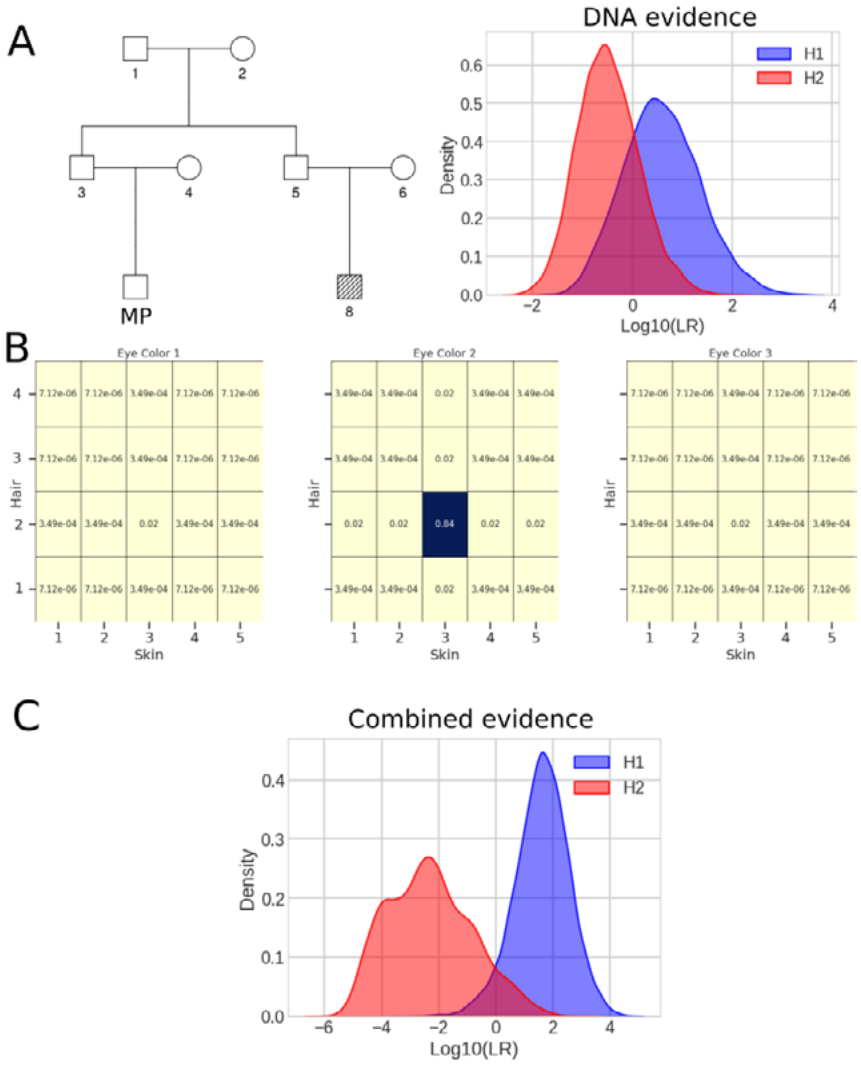


Figure 1. A. LR distributions considering H1 and H2 as true for a case where just one cousin is available for genotyping. **B.** Probability distributions of non-genetic traits under hypotheses H_1 (UP is MP) The probability for a complete match of characteristics between UP and MP is denoted in blue. Color codes are the following - Hair: blonde (1), brown (2), black (3), red (4), Skin: from 1 being the palest to 5 the darkest, Eye: black (1), brown (2), green-blue (3). **C.** LR distributions resulted from integrating genetic and non-genetic LRs. This enhances the separation between the LR distributions under the two hypotheses, reducing the chances of false positives and false negatives.

4. Discussion

In this study, we applied mathematical models recently developed [4, 5] to compute the statistical weight of evidence for non-genetic variables, specifically focusing on pigmentation traits like hair, skin, and eye color. By integrating these non-genetic characteristics with genetic data, we aimed to improve the identification process in missing person investigations, especially in cases where genetic information is limited or underpowered and large scale database searches are carried out. Our results demonstrated that combining genetic and pigmentation trait evidence significantly improves the discrimination between matching and non-matching individuals. This integrated approach provides a more robust framework for evidence evaluation, moving beyond previous methods by accounting for the conditional dependencies among pigmentation traits and avoiding filtering strategies that do not consider the possibility of errors, such as data entry or interpretation problems [4].

However, several limitations need to be addressed. The reliability of the models heavily depends on the availability and accuracy of reference population databases for pigmentation traits, which can vary significantly across different biogeographic ancestries. Additionally, our assumption of independence between genetic and non-genetic evidence may not hold true in all cases, as both types of data can be correlated due to shared ancestral backgrounds. Error rates for non-genetic data sources, such as testimonies or DNA phenotyping methods, also introduce uncertainties that can impact on the reliability of the LR calculations.In many cases such error rates are difficult to estimate or specify. Addressing these challenges requires further refinement of the models, incorporation of more comprehensive population data, and consideration of potential dependencies between evidence types. Future work should focus on expanding these models to accommodate a wider range of forensic scenarios and on developing guidelines for their practical application in real-world investigations.

5. Conclusion

This paper applies recently developed models for calculating likelihood ratios using a set of phenotypic variables, particularly pigmentation traits. Moreover, we have shown the impact of incorporating these likelihood ratios in database search performance through computational simulations. The calculation algorithm and functions are implemented in mispitools, freely available on CRAN.

6. Conflict of interest statement

The authors declare no conflict of interest.

7. References

[1] Puerto, M. S., Abboud, D., Baraybar, J. P., Carracedo, A., Fonseca, S., Goodwin, W.,... & Finegan, O. (2021). The search process: Integrating the investigation and identification of missing and unidentified persons. Forensic science international: synergy, 3, 100154.

[2] Katsara, M. A., Branicki, W., Pośpiech, E., Hysi, P., Walsh, S., Kayser, M.,... & VISAGE Consortium. (2021). Testing the impact of trait prevalence priors in Bayesian-based genetic prediction modeling of human appearance traits. *Forensic Science International: Genetics*, *50*, 102412.

[3] Marsico, F. L., Vigeland, M. D., Egeland, T., & Piñero, M. H. (2021). Making decisions in missing person identification cases with low statistical power. *Forensic science international: genetics*, *54*, 102519.

[4] Marsico, F. L., & Caridi, I. (2023). Incorporating non-genetic evidence in large scale missing person searches: A general approach beyond filtering. *Forensic Science International: Genetics*, *66*, 102891.

[5] Marsico, F., & Egeland, T. (2024). Likelihood Ratios for physical traits in forensic investigations. *bioRxiv*, 2024-05.

[6] Marsico, F. L. (2024). mispitools: An R package for comprehensive statistical methods in Kinship Inference. *bioRxiv*, 2024-08.

[7] Vigeland, M. D., Marsico, F. L., Pinero, M. H., & Egeland, T. (2020). Prioritising family members for genotyping in missing person cases: a general approach combining the statistical power of exclusion and inclusion. *Forensic Science International: Genetics*, *49*, 102376.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 475-482
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Informativity offered by 12 X-STRs (Argus X-12 kit) and 7-XSTRs (ForenSeq DNA Signature prep) from analysis of real kinship cases

Mayra Elizabeth García-Aceves[1], Irán Cortés-Trujillo[1], Nayeli Lizbeth Bernal-Avalos[1], José Alonso Aguilar-Velázquez[2], Héctor Rangel-Villalobos[1]

[1]*Instituto de Investigación en Genética Molecular, Universidad de Guadalajara (CUCiénega-UdeG), Ocotlán, Jalisco México*

[2]*Laboratorio de Ciencias Morfológico-Forenses y Medicina Molecular, Departamento de Morfología, Centro Universitario de Ciencias de la Salud, Guadalajara, Jalisco, México.*

** Héctor Rangel Villalobos, Instituto de Investigación en Genética Molecular, Universidad de Guadalajara (CUCiénega-UdeG), Av Universidad 1115, Paso Blanco, 47810 Ocotlán, Jalisco México, hrangel13@hotmail.com, Tel: +52(392)1000979; fax +52(392)9257112*

Abstract

The informativity a posteriori from real kinship cases offers a worthy landscape for paternity testing casework. This is important for auxiliary and more recently implemented genetic systems, such as the Investigator ArgusX-12 system (Qiagen) and the 7-XSTRs analyzed by MPS of the ForenSeq™ DNA Signature Prep Kit (Verogen). The 7 X-STRs of the ForenSeq kit are a subset of the ArgusX-12 kit; thus, most of the available population databases for Argus X-12 (https://famlink.se/fx_download.html) can be used to create databases for the interpretation of 7 X-STR profiles from Forenseq kit, but at capillary electrophoresis (CE) level. This study aims to describe the a posteriori informativity from complex kinship cases and to compare 12 X-STRs and 7 X-STRs. For this purpose, we analyzed 24 complex kinship cases including 11 grandmother-granddaughters, eight full sisters, and five paternal half-sisters. Exact likelihood ratios (LRs) were estimated with an updated version of the Mexican database (n=1115) for the FamlinkX software, which were created for 12 X-STRs and 7 X-STRs. For each type of kinship case, X-STR set, and linkage group (LG), we estimated the average and range of LR values. For some X-STR loci, some extreme LR values (large and discrete) were observed in different cases, particularly between full sisters (e.g. LR-DXS10101= 1.2E+03), which increased the average LR for these cases (LR= 7.0E+04). Based on 12 X-STRs, the LR was larger for full-sister cases (5.1E+11), followed by grandmother-granddaughter (1.1E+08), and paternal half-sisters (6.4E+04). Conversely, considering an LR >1000 (W ≥ 99.9%) as a conclusive result, full sisters showed the lowest success

rate 75% (6/8), whereas 100% of the paternal half-sister and grandmother-granddaughter cases were conclusive. On the other hand, a considerable LR decrease was observed using only 7 X-STRs concerning 12 X-STRs ($p>0.05$). This is evident with the low success rate based on conclusive results (average LR= 38%): 37.5% (3/8) for full sisters, 40% (2/5) for paternal half-sisters, and 36.4% (4/11) for grandmother-granddaughter cases. In brief, the largest LR values were observed in full-sister cases, followed by paternal half-sister, and grandmother-granddaughter cases. Conversely, with 7 X-STRs, LR values and the proportion of conclusive results diminished significantly.

Keywords
ForenSeq, X-STRs, Kinship cases, FamlinkX, LR

Introduction

Due to their peculiar inheritance pattern, short tandem repeats in the X chromosome (X-STRs) allow obtaining conclusive results of some complex kinship cases where autosomal STRs use to fail, such as those involving half-siblings or grandmother-granddaughter. Among the available X-STR systems, the usefulness of the Investigator Argus X-12 system (Qiagen) has been demonstrated in real complex kinship cases [1]. For this reason, massively parallel sequencing (MPS) platforms have incorporated X-STRs into forensic practice, such as the ForenSeq™ DNA Signature Prep Kit (Verogen). Particularly, the ForenSeq kit analyzes seven of the 12 X-STRs of the Argus X-12 kit, including three two-loci linkage groups (LG): DXS10135-DXS8378 (LG1), DXS7131-DXS10074 (LG2), DXS10103-HPRT (LG3), plus DXS7423 of the LG4 [2].

Although the informativity *a posteriori* offered by ForenSeq kit in real forensic cases has been scarcely reported [3,4], there is no data about the forensic performance of their 7-X-STRs. In fact, the Forenseq ability to solve pairwise 2nd-degree kinships has been evaluated, but X-STRs were excluded from the analysis [3]. This is probably due to the scarce worldwide population databases for this MPS kit, which contrasts with the number of ArgusX-12 population databases (https://famlink.se/fx_download.html). Interestingly, because X-STR profiles of these two kits are compatible at the capillary electrophoresis (CE) level, it is possible to create robust length-based (LB) population databases for these 7 X-STRs from those available for the ArgusX-12 kit. These provisional databases will allow immediate interpretation of forensic cases involving 7 X-STRs, until robust sequence-based (SB) population databases from MPS are available. For this purpose, we updated the Mexican population database for ArgusX-12 kit (from 933 to 1115), and we created a subset for the 7 X-STRs of the Forenseq kit (Cortes-Trujillo et al., *in prep*).

We aimed to assess the forensic performance of 7 and 12 X-STRs by means of the *a posteriori* informativity obtained from the biostatistical interpretation of three sets of complex kinship cases commonly solved with X-STRs.

Material studied, methods, techniques

Sample

We included 24 complex kinship cases: 11 grandmother-granddaughter, 8 full sisters, and 5 paternal half-sisters. Participants signed an informed consent letter approved by the Ethical Research Committee at the Instituto de Investigación de Genética Molecular of the CUCiénega, University of Guadalajara (CUCienega-UdeG). This study follows the regulations of the General Health Law on research involving human subjects (Mexico) and the Helsinki Declaration. The anonymity of the volunteers was preserved at all times.

Genotyping

DNA was obtained from blood stain samples on FTA paper or saliva swabs, after washing with the Whatman® FTA® Purification Reagent. The PCR amplification was carried out with the Investigator Argus X-12 QS kit (Qiagen, Hilden, Germany), followed by capillary electrophoresis (CE) in an ABI-Prism 3130 genetic analyzer (Applied Biosystems®, Foster City, USA) according to the manufacturer recommendations. For allele calling, we used the software Genemapper version 3.2 (Applied Biosystems® Foster City, USA) helped by the corresponding allelic ladder. Our laboratory participates in the quality control proficiency test annually organized by the Spanish and Portuguese-Speaking Working Group of the International Society for Forensic Genetics (www.ghep-isfg.org), with the Argus X-12 QS system.

Statistical analysis

Exact LRs (Likelihood ratios) were estimated with the FamlinkX software [5], using the default lambda value of the database (l= 1000). We used an updated version of the Mexican database for 12-XSTRs (from 933 to 1155) [6] (Cortés-Trujillo et al. in prep), and a subset of this database created for the 7 X-STRs of the Forenseq kit. For each type of kinship case, we estimated the range and average LR by X-STR, linkage group (LG), and X-STR system. Finally, taking as reference LR= 1000 as a conclusive result, we calculated the proportion of conclusive results by type of kinship case, both for 12 X-STRs and 7 X-STRs.

Results

The exact LRs by different sets of kinship cases is presented in Table 1. Based on 12 X-STRs (Argus X-12), the average of combined LR was larger for full-sister cases (5.1E+11), followed by grandmother-granddaughter (1.1E+08), and paternal half-sisters (6.4E+04). On the other hand, based on 7-XSTRs (Forenseq), the LR reduced drastically in all these complex kinship cases. The highest average LR was observed in the full-sister cases (7.0E+04), followed by grandmother-granddaughter (5.9E+03), and paternal half-sisters (1.5E+03).

The highest combined LRs in the full-sister cases result of extreme LR values at individual loci: DXS10134 (LR= 599) and DXS10134 (LR= 3320). These findings can be explained because these full-sisters shared the same genotypes for both loci (DXS10142: 29,36; DXS10134: 21,23.1), and because alleles with the lowest frequency were implied (0.0017 for alleles 36 and 23.1, respectively).

Considering an LR ≥1000 as a conclusive result (W ≥ 99.9%), we estimated the percentage of conclusive cases obtained with 12 XSTRs and 7 X-STRs. For 12 X-STRs, *full sisters showed the lowest success rate 6/8 (75%) by the presence of two inconclusive cases* (LR= 84.56 and 165.46). On the other hand, *100% of the paternal half-sister and grandmother-granddaughter cases reached conclusive results with 12-XSTRs (Table 1).*

Finally, for 7-XSTRs a low success rate based on conclusive results was observed: 37.5% (3/8) for full sisters, 40% (2/5) for paternal half-sisters, and 36.4% (4/11) for grandmother-granddaughter cases.

Discussion

In this work, we report the *a posteriori* informativeness of two X-STR kits: Investigator Argus X-12 and 7-XSTRs included in the ForenSeq™ DNA Signature Prep Kit. Recently it was reported some complex kinship forensic cases solved with X-STRs in Colombia [7]. However, to our knowledge, this is the first report assessing the forensic performance of two X-STR kits to solve some three sets of complex kinship cases (n= 24), particularly because one X-STR set is analyzed by MPS. In addition, to our knowledge, this is the first time that one LB population database is used for interpreting real kinship cases that -theoretically- can be originated from a forensic genomics kit. The provisional nature of this proposal must be stated, because the sequence based (SB) allele diversity for these 7 X-STRs has been reported

that increase to 66% compared to LB alleles in Mexican population [8]. Therefore, the proper forensic assessing of SB alleles for interpreting kinship cases will be done until robust population databases are available.

Both X-STR kits showed the same LR pattern of among kinship cases; the largest average LR was consistently observed among full-sister cases, followed by grandmother-granddaughter, and paternal half-sisters, respectively. However, among these kinship cases, half-sisters displayed -unexpectedly- the lowest conclusive results rate with 12 X-STRs (75%) (Table 1). This finding is explained by the presence of high frequency alleles and haplotypes shared between the alleged relatives. This is worthy because advice to geneticists about the possibility of obtain no conclusive results in full-sister cases with 12 X-STRs, and the possible necessity to add more markers to reach robust results [9,10]. We also must state about the arbitrary criterion to define "conclusive results", because LR= 10,000 (W= 99.99%) could also be taken into account, which would modify our discussion. In brief, a high rate of conclusive results was observed in these complex kinship cases from 12 X-STRs (91.6%) (Table 1), which confirms its convenience due to relatives share IBD (identical-by-descent) alleles leading to powerful statistical results [11].

As could be expected, decreasing the number of loci from 12 X-STRs to 7 X-STRs diminished the combined informativeness in all kinship cases (Table 1). This reduction was critical for the linkage group 4 (LG4), which diminished from three to only one X-STR (DXS7423). Although with 7-XSTRs a drastic LR decrease was observed in all cases (Table 1), and the proportion of conclusive results was limited (38%). These results highlight the complementary nature of this set of 7 X-STRs regarding the remaining loci included in the primer set A of the ForenSeq™: 27 autosomal STRs, 24 Y-chromosomal STRs and 94 identity-informative SNPs [12].

Conclusion

We evaluated the forensic performance on 12 X-STRs and 7 X-STRs, the last obtained from MPS, by means of evaluation of three sets of complex kinship cases. We demonstrated how a population database of LB alleles can be used for this purpose. For both STR kits, the full-sister cases showed the largest LR values, followed by grandmother-granddaughter and paternal half-sister cases. A large proportion of conclusive results was observed with 12-X-STRs, but this proportion diminished significantly with 7 X-STRs.

Acknowledgments

We are grateful with Mexican families who kindly accepted to participate in this study. We thank to CONAHCYT for the posdoctoral scholarship granted to M.E.G-A.

Conflict of interest statement

Authors declare no conflicts of interest.

References

[1] M.Y. Chen, C.W. Ho, C.E. Pu, &, F.C. Wu, Genetic polymorphisms of 12 X-chromosomal STR loci in Taiwanese individuals and likelihood ratio calculations applied to case studies of blood relationships, *Electrophoresis*. 35,12-13 (2014) 1912-20. doi.org/10.1002/elps.201300645.

[2] C. Xavier, W. Parson, Evaluation of the Illumina ForenSeq™ DNA Signature Prep Kit - MPS forensic application for the MiSeq FGx™ benchtop sequencer, Forensic *Sci Int Genet*. 28 (2017)188-194. doi.org/10.1016/j.fsigen.2017.02.018

[3] M. Xu, Q. Du, G. Ma, et al. Utility of ForenSeq™ DNA Signature Prep Kit in the research of pairwise 2nd-degree kinship identification, Int *J Legal Med*. 133(6) (2019) 1641-1650. doi.org/10.1007/s00414-019-02003-6

[4] E. Pilli, R. Tarallo, P. Riccia, A. Berti, A. Novelletto, Kinship assignment with the ForenSeq™ DNA Signature Prep Kit: Sources of error in simulated and real cases, Sci *Justice*. 62(1) (2022) 1-9. doi.org/10.1016/j.scijus.2021.10.007.

[5] D. Kling, B. Dell'Amico, AO. Tillmar, FamLinkX–implementation of a general model for likelihood computations for X-chromosomal marker data, *Forensic Sci Int Genet*. 17 (2015) 1-7. doi.org/10.1016/j.fsigen.2015.02.007

[6] I. Cortés-Trujillo, F. Zuñiga-Chiquette, B. Ramos-González et al. Allele and haplotype frequencies of 12 X-STRs in Mexican population, *Forensic Sci Int Genet*. 38 (2019) e11-e13. doi.org/10.1016/j.fsigen.2018.10.012.

[7] D. Quiceno Cerinza, N. Motta Martín, S.A. Carrillo, C.A. Mora-Torres, O.J. Romero-García, A.P. Robles-Nieto, M. Herrera-Piñero, R.P. Lizarazo-Quintero. Application of the x-chromosomal STRs and the FamLinkX software for solving complex kinship forensic genetics cases in Colombia, *Int J Forensic Sci*. 9 (2024) 000347. doi.org/10.23880/ijfsc-16000347

[8] J.A. Aguilar-Velázquez, M.A. Duran-Salazar, M.F. Córdoba-Mercado, et al. Characterization of 58 STRs and 94 SNPs with the ForenSeq™ DNA signature prep kit in Mexican-Mestizos from the Monterrey city (Northeast, Mexico), *Mol Biol Rep*. 49(8) (2022) 7601-7609. doi.org/10.1007/s11033-022-07575-y

[9] A.O. Tillmar, D. Kling, J.M. Butler, W. Parson, M. Prinz, P.M. Schneider, T. Egeland, L. Gusmão, DNA commission of the International Society for Forensic Genetics (ISFG): Guidelines on the use of X-STRs in kinship analysis, Forensic Sci Int Genet 29 (2017) 269-275. https://doi.org/10.1016/j.fsigen.2017.05.005.

[10] Zhang Y, Yu Z, Mo X, et al. Comparative evaluation of autosomal STRs and X-chromosome STRs as a complement of autosomal STRs in kinship testing in Southern Han Chinese. *Ann Hum Biol*. 2021;48(1):66-69. doi.org/10.1080/03014460.2020.1856926

[11] I. Gomes, N. Pinto, S. Antão-Sousa, V. Gomes, L. Gusmão, A. Amorim. Twenty years later: A comprehensive review of the X chromosome use in forensic genetics. *Front Genet*. 2020;11:926. doi.org/10.3389/fgene.2020.00926.

[12] Verogen. MiSeq FGx™ Instrument Reference Guide. San Diego, CA 92121, USA. 2018.

Table 1. LR values from three kind of complex kinship cases analyzed with 12 and 7 X-STRs, including range and average among loci, linkage group (LG), and combined exact LR, plus the proportion of conclusive results (LR >1000).

Kinship case		Range		Average					
12 X-STRs	Average	LR maximum	LR minimum	LG 1	LG 2	LG 3	LG 4	Combined exact LR	LR> 1000 (%)
Full-sisters	51.03	3.32E+03	0.18	6.2E+04	5.1E+02	1.7E+02	1.7E+04	**5.1E+11**	6/8 (75%)
Paternal half-sisters	4.29	46.46	0.01	2.7E+01	3.3E+01	9.6E+00	2.5E+02	**6.4E+04**	5/5 (100%)
Grandmother-granddaughter	8.69	360.31	0.24	1.7E+02	3.9E+02	3.2E+01	8.7E+01	**1.1E+08**	11/11 (100%)
Kinship case		Range		Average					
7 X-STRs	Average	LR maximum	LR minimum	LG 1	LG 2	LG 3	DXS7423	LR exact	LR> 1000 (%)
Full-sisters	25.17	1.2+03	0.3	3.6E+02	2.5E+01	8.5E+00	2.4E+00	**7.0E+04**	3/8 (37.5%)
Paternal half-sisters	4.0	29.7	0.1	2.9E+01	2.1E+01	3.4E+00	3.6E+00	**1.5E+03**	2/5 (40%)
Grandmother-granddaughter	11.6	401.5	0.2	1.0E+02	1.3E+02	6.1E+00	2.8E+00	**5.9E+03**	4/11 (36.4%)

The corresponding FamlinkX software database was employed for 12 X-STRs or 7 X-STRs.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 483-489
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Exploring different software options for likelihood ratio calculations in paternity cases with low template DNA

Alessia Riem[1*], Elena Chierto[1], Serena Aneli[1],
Federica Bertolotto[1], Carlo Robino[1]

[1] *Department of Public Health Sciences and Pediatrics, University of Turin, Turin, Italy.*

** Corresponding author at: alessia.riem@unito.it, (+39)0116705625*

Abstract

In low template DNA (LT-DNA) samples, the interpretation of STR profiles by calculation of likelihood ratio (LR) must consider the possibility of stochastic artefacts. Several probabilistic genotyping software were developed for this purpose with main focus on the direct identification of stain contributors. Less attention has been given to kinship cases requiring the analysis of LT-DNA specimens.

We evaluated the performances of three software -Familias, EuroForMix (EFM) and EFMrep- in 25 paternity cases that required the analysis of LT-DNA samples, including formalin-fixed, paraffin-embedded archival tissues, bone specimens, and personal items. Tested pedigrees consisted of both trios and duos (missing maternal information). LT-DNA samples were classified as "highly" (HD) or "mildly" (MD) degraded, based on the quality of STR profiles (risk of drop-out, estimated through a logistic model). Calculations were performed using two PCR results from each LT-DNA sample (two PCR replicates of the same STR kit for Familias and EFM, and two different STR kits for EFMrep).

It was observed that in MD-duo and HD-duo cases average LR values obtained with EFMrep were significantly higher compared to Familias and EFM. When maternal information was available (trios), Familias generated significantly higher LRs compared to EFM and EFMrep (MD samples), whereas Familias and EFMrep were equivalent in maximizing LR values in HD samples. However, the proportion of HD-trio cases with LR values passing a 10,000 threshold was higher for EFMrep (60%) compared to Familias (40%)

These preliminary results can represent the basis for the development of a standardized analytical and interpretative flow chart in kinship cases involving LT-DNA samples.

Keywords

Kinship testing, LT-DNA, probabilistic genotyping.

Introduction

In the last 15 years probabilistic genotyping software has seen a remarkable development, with a special focus on personal identification of low template DNA (LT-DNA) casework samples. Application of such software, like EuroForMix (EFM) [1]the peak height properties, stutter proportion and degradation. In addition, EuroForMix includes models for allele drop-out, allele drop-in and sub-population structure. EuroForMix supports two inference approaches for likelihood ratio calculations. The first approach uses maximum likelihood estimation of the unknown parameters. The second approach is Bayesian based which requires prior distributions to be specified for the parameters involved. The user may specify any number of known and unknown contributors in the model, however we find that there is a practical computing time limit which restricts the model to a maximum of four unknown contributors. EuroForMix is the first freely open source, continuous model (accommodating peak height, stutter, drop-in, drop-out, population substructure and degradation or EFMrep [2], to LT-DNA samples from paternity cases has potentialities and limitations, compared to standard paternity analysis software (e.g. Familias [3]) (Table 1). EFM and EFMrep can deal with quantitative data and the whole spectrum of PCR artefacts related to LT-DNA. On the other hand, Familias considers mutation and takes advantage of maternal information in trio cases. The three software also differ regarding the possibility of accommodating PCR replicate data, with Familias only processing profiles obtained through consensus. With this in mind, we evaluated the performances of Familias, EFM and EFMrep in a set of paternity cases that required the analysis of LT-DNA samples.

Table 1: Strengths and limitations in the use of the three different software considered for this study in presence of LT-DNA paternity cases.

	Familias	EFM	EFMrep
Drop out	Yes	Yes	Yes
Drop in	No	Yes	Yes
Stutter	No	Yes	Yes
Mutation	Yes	No	No
Maternal information	Yes	No	No
Peak height	No	Yes	Yes
PCR replicas (single STR kit)	No (consensus)	Yes	Yes
PCR replicas (multiple STR kits)	No (consensus)	No	Yes

Material studied, methods, techniques

Samples

25 paternity cases each including a LT-DNA sample were selected: formalin-fixed, paraffin-embedded (FFPE) archival tissue (28%), bone specimens (52%), personal items (20%). Of these, 68% of the cases were trios and 32% were duos (motherless cases).

Genotyping

Samples were amplified with the PowerPlex ESI 17 Fast (Promega) (ESI) and/or PowerPlex ESX 17 Fast (Promega) (ESX) STR kits; PCR products were run on the SeqStudio Genetic Analyzer and genotyped with GeneMapper software v 5.1 (ThermoFisher Scientific).

LT-DNA samples were classified as highly (HD) or mildly (MD) degraded depending on bp length below/above 200 bp of the longest STR amplicon with peak height above stochastic threshold associated to < 5% risk of drop out determined through a logistic model [4]and thus different concentrations, in the presence of degradation. The possibility of deriving information on DNA degradation was evaluated in a forensic qPCR assay not specifically designed to detect DNA fragmentation, the Plexor HY (Promega.

Biostatistical analysis

Likelihood ratio (LR) calculations were performed using:

— Allelic frequencies observed in the Italian population [5,6]except for locus SE33 [7]
— Mutation rates reported at https://strbase.nist.gov/
— F_{ST} =0.01 [8]
— For EFM/EFMrep, drop-in and lambda parameters calculated with the EFM "Fit-in drop-in data" function based on internal validation data

Familias (version 3.3.1)

From 2 PCR replicates of a single STR kit (ESI or ESX) a "consensus" profile was obtained for the LT-DNA sample; in loci where only a single allele was present above the stochastic threshold in both replicates, the genotype was assumed to be homozygous with drop out probability calculated according to logistic regression [9]. In trio cases calculations were done both using and disregarding maternal information.

EFM (version 4.0.8)

LT-DNA samples underwent 2 PCR replicates with a single STR kit (ESI or ESX). The two alternative hypotheses used in the calculations were:

— Hp: sample comes from an unknown individual;

— Hd: sample comes from the parent/child of the person for which a reference (non-LT-DNA) sample was available in the paternity case.

The reciprocal of the obtained LR value was then used for comparisons between software.

EFMrep (version 1.1.0)

Each LT-DNA sample was subjected to amplification using two different STR kits (ESI and ESX). The two alternative hypotheses were as in EFM, but with inverted numerator/denominator.

Statistical analysis

LRs were compared by t- or Wilcoxon test depending on normality (assessed with Shapiro-Wilk test) using *jamovi* (Version 2.5) software.

Results

LR calculation results are summarized in Figure 1.

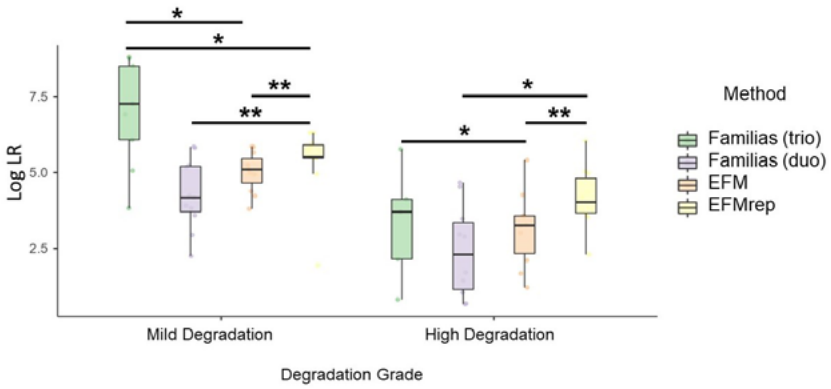


Figure 1. Boxplot showing LR values calculated by each software divided by degree of degradation of LT-DNA samples. Familias LR values were obtained either using maternal information when available ("trio") or disregarding it ("duo").

It can be seen that, in duo cases, EFMrep LR values were significantly higher than those of Familias (HD, $p < 0.05$; MD, $p < 0.01$) and EFM (HD and MD, $p < 0.01$). No significant difference existed between Familias and EFM.

When maternal information was available (trios), Familias LR values were significantly higher than those of EFM (HD and MD, $p < 0.05$). Familias also showed significantly higher LRs than EFMrep ($p < 0.05$) in MD samples. In HD-trio samples LR results of Familias and EFMrep did not differ significantly.

The proportion of cases passing the LR > 10,000 paternity threshold recommended by current Italian guidelines (http://www.gefi-isfg.org/temp/20112018100854.pdf), depending on the type of pedigree, DNA degradation status and type of software used for calculations are shown in Table 2.

Table 2. Proportion of paternity tests with final LR value > 10,000.

	Familias	EFM	EFMrep
HD-duo	20%	20%	50%
MD-duo	55%	91%	91%
HD-trio	40%	20%	60%
MD-trio	89%	89%	89%

It can be seen that, despite the higher average LR values obtained with Familias in MD-trios, the proportion of cases reaching the paternity threshold of 10,000 did not differ between the three software (89%). On the other hand, although Familias and EFMrep average LRs were comparable in HD-trios, the proportion of cases showing LR values > 10,000 was higher for EFMrep (60%) compared to Familias (40%).

Discussion

There is a current shortage of dedicated probabilistic genotyping software for kinship testing of LT-DNA samples [10]for example determining whether two persons are second cousins or unrelated, can be done by comparing their genotypes at a selection of genetic markers. When the data for one or more of the persons is from low-coverage next generation sequencing (lcNGS. Standard software, such as Familias, can model dropout. However, it lacks the whole range of functionalities associated with continuous probabilistic software used for the interpretation of individual contributions to stains. We took advantage of the ability of such software (EFM, EFMrep) to deal with hypotheses including related contributors to test their performance in paternity cases with LT-DNA, as an alternative to Familias. The obtained results indicate that, independently from the typology of pedigree (with or

without maternal information) and from the degree of degradation of the sample, EFMrep was the software that maximized LR values, due to its ability to analyse data derived from the combination of STR kits with different primer design, thus increasing the number of fully informative loci [2]. The only exception was seen in cases where DNA degradation was only mild and maternal information was available. In such conditions, standard analysis with Familias generated significantly higher LR values compared to EFM and EFMrep, but no difference between the three software in the percentage of cases passing a set paternity threshold of LR > 10,000 was seen.

Conclusion

This preliminary study enabled us to define a flow chart for the selection of preferential amplification and interpretation strategies in paternity cases that may require the analysis of LT-DNA, such as that derived from FFPE tissue, bone specimens, or personal items used as reference samples. To strengthen current findings, analysis of further LT-DNA paternity cases and investigation of the impact on software performance of additional replicates and combinations of STR kits will be needed in the future.

Acknowledgments

None.

Conflict of interest statement

None.

References

[1] Ø. Bleka, G. Storvik, P. Gill, EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts, Forensic Sci Int Genet 21 (2016) 35–44.

[2] Ø. Bleka, L. Prieto, P. Gill, EFMrep: An extension of EuroForMix for improved combination of STR DNA mixture profiles, Forensic Sci Int Genet 61 (2022) 102771.

[3] D. Kling, A.O. Tillmar, T. Egeland, Familias 3 - Extensions and new functionality, Forensic Sci Int Genet 13 (2014) 121–127.

[4] E. Chierto, S. Aneli, N. Nocco, et al, Assessing DNA Degradation through Differential Amplification Efficiency of Total Human and Human Male DNA in a Forensic qPCR Assay, Genes (Basel) 15 (2024) 622.

[5] S. Presciuttini, N. Cerri, S. Turrina, et al, Validation of a large Italian Database of 15 STR loci, Forensic Science International 156 (2006) 266–268.

[6] C. Previderè, P. Grignani, F. Alessandrini, et al, The 2011 GeFI collaborative exercise. Concordance study, proficiency testing and Italian population data on the new ENFSI/EDNAP loci D1S1656, D2S441, D10S1248, D12S391, D22S1045, Forensic Sci Int Genet 7 (2013) e15-18.

[7] L.A. Welch, P. Gill, C. Phillips,et al, European Network of Forensic Science Institutes (ENFSI): Evaluation of new commercial STR multiplexes that include the European Standard Set (ESS) of markers, Forensic Sci Int Genet 6 (2012) 819–826.

[8] J. Buckleton, J. Curran, J. Goudet, et al Population-specific FST values for forensic STR markers: A worldwide survey, Forensic Sci Int Genet 23 (2016) 91–100.

[9] P. Gill, L. Gusmão, H. Haned, et al, DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods, Forensic Science International: Genetics 6 (2012) 679–688.

[10] P. Mostad, A. Tillmar, D. Kling, Improved computations for relationship inference using low-coverage sequencing data, BMC Bioinformatics 24 (2023) 90.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 491-498
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Ge.F.I. updated recommendations on the collection of biological samples for forensic genetic testing in the medical legal care of victims of sexual violence and/or abuse

Susi Pelotti[1*], Federica Alessandrini[2], Filippo Barni[3], Andrea Berti[4], Carla Bini[1], Loredana Buscemi[2], Sarah Gino[5], Fabio Oldoni[6], Andrea Piccinini[7], Carlo Robino[8], Paolo Fattorini[9]

[1]*Section of Legal Medicine, Department of Medical and Surgical Sciences, Alma Mater Studiorum, University of Bologna, Bologna, Italy*

[2]*Section of Legal Medicine, Department of Biomedical Sciences and Public Health, Polytechnic University of Marche, Ancona, Italy.*

[3]*Carabinieri Scientific Investigation Department of Roma, Italy*

[4]*Carabinieri Scientific Investigation Department of Cagliari, Italy*

[5]*Department of Health Sciences, University of Eastern Piedmont, Direzione Medica dei Presidi Ospedalieri, AOU Maggiore della Carità, Novara, Italy*

[6]*Department of Chemistry and Physics, Arcadia University, Pennsylvania, USA*

[7]*Department of Biomedical Sciences for Health, University of Milan, Milan, Italy*

[8]*Department of Public Health Sciences and Pediatrics, University of Turin, Turin, Italy*

[9]*Department of Medicine, Surgery and Health, University of Trieste, Ospedale di Cattinara, Trieste, Italy*

** Via Irnerio 49, Bologna, Italy, 0039 (0)51 2088343, susi.pelotti@unibo.it*

Abstract

The purpose of the Ge.F.I. recommendations update is to guide the intervention of healthcare professionals involved in the collection of biological samples from victims of sexual violence and/or abuse for forensic genetic testing purposes, considering the new DNA technologies and the increase of analysis sensitivity. The sample collection procedure is crucial for ensuring that biological stains from victims of sexual violence are properly and effectively collected to avoid DNA contamination or degradation as well as the assessment of the chain of custody to maintain evidence integrity.

The recommendations serve as the "minimum standards" that can be included in the practices of healthcare providers and response units to sexual violence and abuse.

Informed consent from victims of sexual violence and/or abuse must be properly recorded before receiving medical care or participating in any body sampling activity. Healthcare professionals—even those appointed by the Judicial Authority— must ensure that individuals seeking assistance in cases of sexual violence and/or abuse receive comprehensive and unambiguous information about the protocols for the collection of biological specimens for forensic genetic testing. Guiding criteria for handling, sampling, and storage of forensic evidence, including samples from body and clothing or other material of interest of the sexual assault victim, are described in a table format, which is easy to consult also in the emergency room. In addition, some photos showing the proper techniques for collection of skin samples and subungual material are included. Finally standard forms for informed consent, sample collection, chain of custody, and a list of materials/supplies useful for the collection of forensic evidence in emergency departments are provided. These forms can greatly assist the documentation process.

Keywords
Sexual violence, biological samples, collection, DNA, chain of custody

1. Introduction

The Ge.F.I., Italian working group of the Internation Society for Forensic Genetics (ISFG), updated the previous published recommendations on the collection of biological samples for forensic genetic testing in the forensic care of victims of sexual violence and/or abuse. The purpose of the Ge.F.I. updated recommendations is to guide the intervention of healthcare professionals considering the new DNA technologies and the increase in their analytical sensitivity. These recommendations have the following main objectives:

- outline standard medical procedures for sampling of biological specimens from survivors of sexual assault and/or abuse and which can be accessible to judicial authority for the proper forensic examination/testing;
- ensure the evidentiary value of biological specimens by properly maintaining the integrity and continuity of the chain of custody for handling, sampling, and storage of forensic evidence;
- provide adequate medical forensic examination procedures in instances involving victims of sexual assault and/or abuse.

2. Material studied, methods, techniques

Members of the Ge.F.I. working group revised the scientific literature and the international recommendations and guidelines. Group meetings were organized to

evaluate each regional healthcare settings to adapt the procedures to the local facilities and health professionals training. Standard forms for informed consent, samples collection, chain of custody, and a list of materials/supplies useful in the collection of forensic evidence in hospital emergency departments were recommended based on the Italian law and code of penal procedure.

3. Results

3.1 Informed consent

Informed consent from victims of sexual violence and/or abuse must be properly recorded before receiving medical care or participating in any body sampling activity. Healthcare professionals—even those appointed by the Judicial Authority—must ensure that individuals seeking assistance in cases of sexual violence and/or abuse receive comprehensive and unambiguous information about the protocols for the collection of biological specimens for forensic genetic testing. They also have to ensure that the person receiving medical care is given proper attention to any question about the reasons and procedures of the sampling process. Both acceptance and refusal of consent should be formally documented in the medical history records. If, after being fully informed, the person does not consent, healthcare professionals should record the victim's disagreement and not proceed further with sampling. For minors, individuals exercising parental responsibility or legal guardianship must provide or refuse their informed consent to any medical treatment the minor receives taking into account the will of the child in light of his/her age and maturity. Medical practitioners are formally required to inform those with parental responsibility or the legal representative that, in the event of a refusal to give consent, the case will be reported to the Judicial Authority if *ex officio* crime prosecution is involved. This notification must be provided after they have been duly informed about the diagnostic hypothesis of violence and/or sexual abuse against the minor and the procedure for sampling on the minor's body sites.

3.2 Samples collection

A detailed examination of the circumstances surrounding the violent event is necessary before collecting any biological specimens. When feasible, it is important to take into account the following factors: the length of time elapsed between the violent event and the sample collection, how the violent act was committed, the actions performed by the survivor after the assault (such as cleaning the affected anatomical parts, changing clothes, ingesting liquids, etc.), and the psychophysical

state of the victim. It is recommended to collect biological specimens from the victim for forensic DNA testing as soon as possible, as the probability of collecting useful evidentiary material decreases over time. It is also advisable to collect samples on the person receiving assistance and obtain clothing and personal items in cases of prolonged mechanical immobility as a result of the victim's impairment, or if there is no specific information available regardless the time since the aggression.

It is preferable to consult a forensic genetics expert whenever possible for guidance on the handling, collection, and preservation of biological evidence given the unique nature of forensic DNA testing. Biological specimens (blood, semen, and saliva stains, hairs, sweat and epidermal cells) can be collected on different body sites including swabs from oral, skin, external genital, vaginal, cervical, perianal, rectal and penile areas. Hairs and sub-ungueal material must be collected.

Other biological specimens could also be found at the place where the violent event occurred: these are special situations that involve the intervention of the Judicial Authority and/or law enforcement agency. The healthcare practitioner responsible for gathering the circumstantial data must notify the Judicial Authority and/or law enforcement agency of this possibility based on the narrative disclosed.

Biological specimens on clothing are not always visible to the naked eye. All clothing worn at the time of the event should therefore be collected, with particular regard to underwear and any intimate pads, diapers, or whatever is in contact with the intimate regions, which, if found attached to the garment, should not be detached. If the person receiving assistance claims to have changed clothes, it is preferable to obtain clothing currently wearing, especially the underwear, and advise to preserve (without washing) the items worn at the time of the violent event to be turned in either directly to law enforcement or indirectly through medical staff.

Items such as handkerchiefs, condoms (including female condoms), and speculum should be collected, placed in appropriately sized containers (e.g., disposable falcons, Petri dishes, jars), and stored in the freezer at -20°C. In the instance of condoms, in particular, it is instructed to seal their end to prevent the leakage of any seminal fluid. It is mandatory to change gloves after this operation. Since it is also possible to generate genetic profiles in the absence of sperm cells from swabs (vaginal, rectal, etc.), setting up microscope slides for sperm detection is not necessary for forensic genetic testing. As microscope slides used for genetic analysis can also be set up for other diagnostic purposes, samples should be stored in suitable containers to avoid contamination and maintain the integrity of the chain of custody.

In case of voluntary termination of pregnancy, miscarriage, or childbirth related to violent events, healthcare personnel (gynecologist/obstetrician) must

collect abortion material, placenta, fetal funiculus, and neonatal blood. Abortion material/placenta/fetal funiculus should be stored in separate and sterile containers as dry as possible, avoiding excess blood, without preservative liquid (e.g., formalin or saline), and frozen at -20°C. In case of childbirth it is recommended to take a few drops of neonatal blood to be affixed on a specific paper (e.g., Guthrie card).

3.3 Protective devices

All healthcare practitioners must mandatory wear personal protective equipment, adopt all procedures aimed at avoiding contamination events (personnel-specimen contamination, specimen-specimen contamination, environment-specimen contamination), and ensure the integrity and proper handling, storage, and traceability of the evidentiary material collected.

A health care provider is responsible for the material collected during the collection stages; thus, it is recommended to avoid leaving the victim unattended at the outpatient clinic while samples are being taken. In the case of injury, such as a bite or wound, it is crucial to discuss with the clinician whether samples of forensic genetic relevance can be obtained before the wound is healed, and if that is possible, proceed with taking metric-referenced photographic images.

Oral swabbing should be considered in the event of oral penetration, and this should be discussed with the clinician as soon as the anamnestic data is obtained and before the victim ingests any oral fluids.

The collected items must be placed in tamper-evident, sealed bags to maintain the integrity of the chain of custody.

It is recommended, according to the internal policies of each healthcare provider facility, to:

1. Draft the collection form and outline all procedures related to the sampling of relevant body samples and clothing/other material of interest;
2. Include the following sentence in the medical records of the sexual assault victim (ER report, medical record, or outpatient certificate): “A collection list including evidence samples has been created in case forensic genetic testing will be required.”
3. Prepare a chain-of-custody form, allowing traceability of the evidence (delivery-custody), which must be countersigned by the healthcare provider responsible of victim care. This record must contain the identifying information of the case, date of the tasks completed, and details of the healthcare provider. The documentation must follow the evidentiary items and must be updated with subsequent handover actions.

Discussion

International guidelines and scientific literature suggest that individuals who have experienced sexual violence often seek medical assistance, even if they may fail to disclose the event itself. As reported by WHO, health workers are also well placed to collect and document the evidence necessary for corroborating the circumstances of the sexual assault, and for identifying the perpetrator and the health consequences of the event. Such evidence is often crucial to the prosecution of sexual violence cases [1]. Since 2013, the Ge.F.I. group published on https://www.gefi-isfg.org the first version of the recommendations on the collection of biological samples for forensic genetic testing in the contest of medical legal care of victims of sexual violence and/or abuse and the document was attached to the Italian National guidelines for health services and hospitals for the social-healthcare assistance to women victims of violence.

Following the international recommendations and dedicated toolkit [2-5], the standard protocol guiding the process of forensic evidence collection has been updated considering the new DNA technologies and their increasing analysis sensitivity. Considering that the sample collection procedure is crucial to avoid DNA contamination or degradation as well as the assessment of the chain of custody to maintain evidence integrity, our recommendations serve as the "minimum standards" that should be included in the practices of healthcare providers and response units to sexual violence and abuse ensuring also the evidentiary value of biological specimens.

For a rapid consultation by health care providers during the physical examination of the persons who report having been sexually assaulted, a quick consultation table was prepared containing the issues on procedures used to collect biological specimens from a sexual assault victim's body, subdivided in: body site of collection, maximum time interval after offense, method of sample collection, and method of intimate sample storage. A summary table on the procedures used to collect the victim's clothing and other material of interest was also set up.

We agree with WHO [1] that health workers should be aware of the capabilities and requirements of their forensic laboratory and we think that that the recommendations can also be used for training of health care practitioners dedicated to preventing and responding to all forms of violence against women and girls. The training, as reported by WHO [6], is crucial to improve the health system's response to violence against women.

Conclusion

The recommendations are available at https://www.gefi-isfg.org and will be incorporated into healthcare practices to provide care and support to survivors of sexual assault *Feedback on these guidelines are welcome. As reported on ISFG website, the Ge.F.I. group is inviting anyone who may wish to contribute to the revision of this document, by sending their comments to the Coordinator of the Sexual Violence Working Group, Susi Pelotti at susi.pelotti@unibo.it.*

Acknowledgments

We are grateful to the Police Bureau of Forensic Services for their assistance in putting this material together.

Conflict of interest statement

None

References

[1] World Health Organization. Guidelines for medico-legal care for victims of sexual violence. World Health Organization, Geneva, 2003. https://iris.who.int/handle/10665/42788

[2] World Health Organization & UNODC. Strengthening the medico-legal response to sexual violence, 2015. https://www.who.int/publications/i/item/WHO-RHR-15.24

[3] Scientific Working Group on DNA Analysis Methods. Recommendations for the efficient DNA processing of sexual assault evidence kits, 2016. https://www.swgdam.org/_files/ugd/4344b0_4daf2bb5512b4e2582f895c4a133a0ed.pdf

[4] The Faculty of Forensic & Legal Medicine of the Royal College of Physicians. Recommendations for the Collection of Forensic Specimens from Complainants and Suspects, 2023. https://fflm.ac.uk/resources/publications/recommendations-for-the-collection-of-forensic-specimens-from-complainants-and-suspects/

[5] National SART Guidelines Development Group. National guidelines on referral and forensic clinical examination following rape and sexual assault (Ireland), 2023. https://assets.hse.ie/media/documents/ncr/Guidelines_on_Referral_following_Rape_and_Sexual_Assault.pdf

[6] World Health Organization. Caring for women subjected to violence: A WHO training curriculum for health care providers. Revised edition, 2021. Geneva, 2021. https://www.who.int/publications/i/item/9789241517102

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 499-504
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Identification of cell death biomarkers in early and late postmortem intervals

Francisco Medina-Paz[1], Sara C. Zapico[1,2*]

[1]*Department of Chemistry and Environmental Sciences, New Jersey Institute of Technology, Newark, NJ.*

[2]*Anthropology Department and Laboratories of Analytical Biology, NMNH, Smithsonian Institution, Washington, DC.*

** Corresponding author at: New Jersey Institute of Technology, Department of Chemistry and Environmental Sciences, 161 Warren St., Tiernan Hall 365, Newark, NJ 07102, USA and Smithsonian Institution, National Museum of Natural History, Anthropology Department and Laboratories of Analytical Biology, 10th and Constitution Ave, NW, PO Box 37012, Washington DC 20560, USA.*
E-mail: sc338@njit.edu and casado-zapicos@si.edu

Abstract

The present work aimed to assess postmortem changes at the cellular level over time, as a potential tool to improve Postmortem Interval (PMI) estimates. To address this, RNA was isolated from buccal swab samples from human subjects of postmortem intervals from 1 to 22 days after death under natural environmental decomposition conditions. Based on the molecular cell death signaling processes, we studied the expression patterns of 18 genes involved in cell death by quantitative Polymerase Chain Reaction (qPCR). Remarkably, we observed expression of both housekeeping (GAPDH), and key genes involved both in apoptosis and necrosis processes up to 22 days after death. Additionally, a constant downregulation of the expression of FasL, FasR, DR4, DR5, and Caspase-9 genes was observed for all timepoints compared to day 1 after death, along with similarity in their expression patterns from 2-10 days after death. Furthermore, Bcl-2, RIPK-1 and NF-κβ genes éxpression was found to be upregulated for all timepoints compared to day 1 after death, together with similar expression patterns from 2 to 10 days after death. Further research is needed to validate these genes as PMI estimation biomarkers by evaluating them in a larger number of individuals and conditions.

Keywords
PMI, biomarkers, cell-death, apoptosis, necroptosis, qPCR.

Introduction

Postmortem interval (PMI) determination is an important and, at the same time, a challenging issue in forensics sciences [1]. Over the years, innovative approaches have been tried, from physical evaluation to molecular methods, without being validated as successful for accurate PMI estimation [2]. Postmortem interval estimation relies on the assessment of the effect of various processes occurring in the body after death. Traditionally, postmortem interval has been classified into early and late PMI. Early PMI (up to 3 days after death) has been most frequently estimated using physical postmortem changes. Late PMI (more than 3 days after death), when the decomposition begins, is a more challenging task as environmental factors and the presence of insects and scavengers can accelerate or slow down these processes; consequently, other methods different to physical evaluation have been developed [3]. The present work aimed to assess postmortem gene expression changes at the cellular level over time as a potential tool to improve PMI estimates. To address this, RNA was isolated from buccal swab samples from human subjects of postmortem intervals from 1 to 22 days after death under natural environmental decomposition conditions.

Material studied, methods, techniques

Sample obtention

The experiment was carried out at the Forensic Osteology Research Station (FOREST) facilities in Cullowhee, North Carolina. Buccal swab samples were taken from human subjects from 1 to 22 days after death under natural environmental decomposition conditions.

RNA isolation and qPCR

DNA/RNA co-isolation was performed using the Quick DNA/RNA Miniprep Plus Kit (Zymo Research. Irvine, CA, USA) from all buccal swab samples using and cDNA was synthesized using the qScript Ultra SuperMix (Quantabio. Beverly, MA, USA) and RT-PCR was conducted using the PerfeCTa SYBR® Green FastMix

(Quantabio. Beverly, MA, USA) for 18 genes involved in cell death processes plus one housekeeping gene (GAPDH), according to manufacturer's recommendations.

Results

Gene expression was found in all samples from 1 to 22 days after death, both for the housekeeping gene and for the rest of the analyzed genes in the present work. During the early PMI (up to 3 days after death) period the expression of genes FasR, FasL, DR4, DR5, TRAIL, and Caspases 3 and 9, all of them involved in the apoptosis pathway, was downregulated compared to day 1 after death. On the other hand, genes PTEN, Akt, Jnk, Bcl-2, Bad, NF-κβ, and Caspase 8, also involved in apoptosis and other regulatory pathways, and the RIPK-1 gene involved in necroptosis, were observed to be upregulated during the early PMI period (Figure 1).

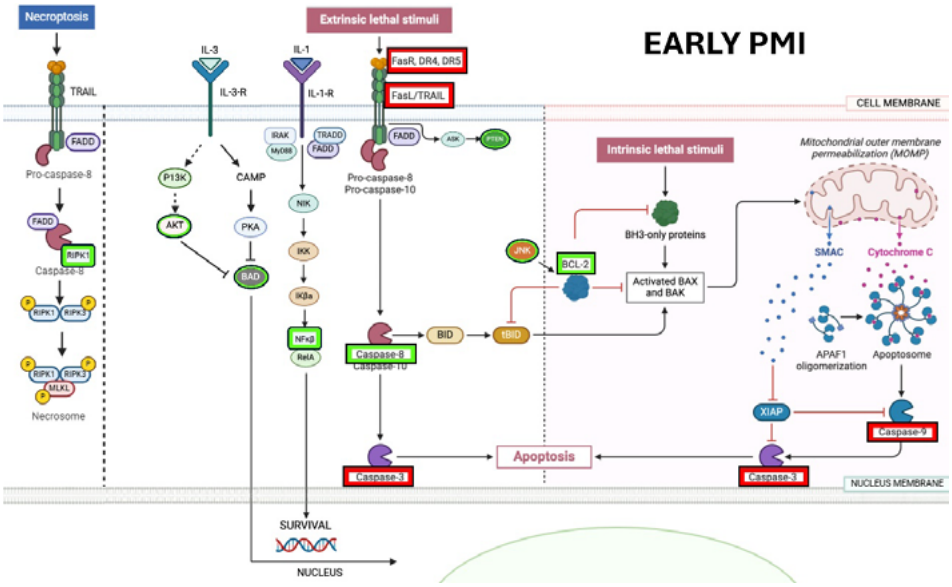


Figure 1. Schematic representation of apoptotic and necroptotic gene expression during early PMI. The genes highlighted in green color represent upregulation, and the genes highlighted in red color represent downregulation during early postmortem interval (PMI).

For late PMI (more than 3 days after death) we observed genes FasR, FasL, DR4, DR5, TRAIL, and Caspases 3 and 9 but also Caspase 8 to present a lower expression compared related to day 1 after death. In regard necroptosis pathway genes, RIPK-1 continued presenting a higher expression in comparison to day 1 after death (Figure 2).

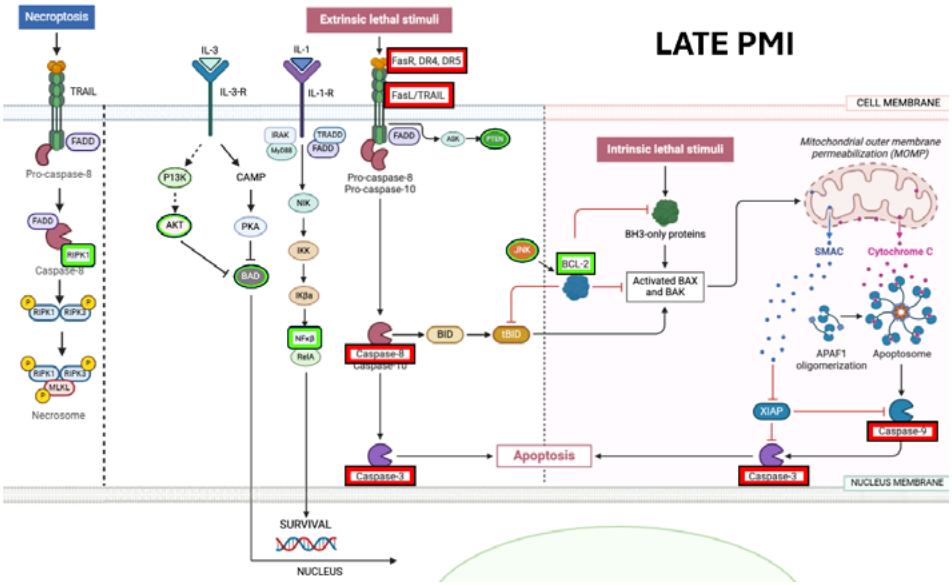


Figure 2. Schematic representation of apoptotic and necroptotic gene expression during late PMI. The genes highlighted in green color represent upregulation, and the genes highlighted in red color represent downregulation during late postmortem interval (PMI).

Discussion

Higher expression, compared to day 1 after death, was found for genes involved in apoptosis inhibition such as Akt (Ak strain transforming), Bcl-2 (B cell lymphoma/leukemia 2), and NF-κβ (Nuclear factor κβ), and both in early and late PMI samples. The expression level of apoptotic inhibitors might be explained as a late attempt of the cell to survive by halting programmed cell death and promoting the activation of survival genes. For example, NF-kB protein has been reported to exert pro-survival activity by inducing the transcription of several antiapoptotic genes (i.e. Bcl-2 family members) in certain tissues under certain biological contexts [4-6]. Some

pro-apoptotic genes: PTEN (Phosphatase and tensin homolog deleted on chromosome 10), Bad (Bcl-2 associated agonist cell death), and Jnk (c-Jun N-terminal kinase), were also observed to be upregulated both in early and late PMI samples. The increase in the expression of genes encoding pro- and antiapoptotic proteins seems to respond to an inter-regulation between them, as it has been observed for other animals such as zebrafish and mice [7].

Furthermore, from the whole set of genes studied in the present work, Caspase-8 was the only one we detected to substantially change its gene expression pattern between early and late postmortem interval as in terms of its expression compared to day 1 after death. Caspase-8 has been reported to trigger the extrinsic apoptotic pathway in response to the activation of cell surface Death Receptors (DRs) such as Fas, TRAIL, and TNF, and it can also trigger the intrinsic apoptotic pathway under certain conditions [8]. Besides, Caspase-8 also plays key roles in autophagy, and pyroptosis as well as playing a crucial function on inhibition of necroptosis [9]. By reducing the expression of the Caspase-8 gene in the late postmortem interval the cell may be making a late attempt to stay alive.

Finally, it is also worth mentioning that during the early and late PMI periods, the expression level of the RIPK1 gene remained higher than on day 1 postmortem, together with the high expression of the anti-apoptotic genes and inhibition of anti-necroptosis/pro-apoptosis Caspase-8 gene in late PMI samples, which might be pointing to the cell moving on from an apoptosis process to a necrosis process.

Conclusion

We observed expression of both housekeeping (GAPDH), and key genes involved both in apoptosis and necrosis pathways up to 22 days after death. A constant downregulation of the expression of FasL, FasR, DR4, DR5, and Caspase-9 genes was observed for all timepoints compared to day 1 after death, along with similarity in their expression patterns from 2-10 days after death; while Bcl-2, RIPK-1 and NF-κβ expression was found to be upregulated for all timepoints compared to day 1 after death, together with similar expression patterns from 2 to 10 days after death. Remarkably, Caspase-8 was the only one we detected to change its gene expression pattern between early and late postmortem interval, upregulated and downregulated, respectively. Further research is needed to validate these genes as PMI estimation biomarkers by evaluating them in a larger number of individuals and conditions.

Acknowledgments

The authors acknowledge Dr. Nicholas V. Passalacqua, Dr. Rebecca George, Autumn Hall and Ryan Bodine, from Western Carolina University for collecting and providing the samples to carry out this study.

Funding source

This project was funded by the New Jersey Institute of Technology Faculty Seed Grants 2022.

Conflict of interest statement

The authors declare no conflict of interest.

References

[1] Henssge, C. and B. Madea, *Estimation of the time since death in the early postmortem period.* Forensic Sci Int, 2004. **144**(2-3): p. 167-75.

[2] Sangwan, A., et al., *Role of molecular techniques in PMI estimation: An update.* J Forensic Leg Med, 2021. **83**: p. 102251.

[3] Madea, B., *Methods for determining time of death.* Forensic Sci Med Pathol, 2016. **12**(4): p. 451-485.

[4] De Smaele, E., et al., *Induction of gadd45β by NF-κB downregulates pro-apoptotic JNK signalling.* Nature, 2001. **414**.

[5] Tang, G., et al., *Inhibition of JNK activation through NF-κB target genes.* Nature, 2001. **414**.

[6] Baldwin, A.S., *Regulation of cell death and autophagy by IKK and NF-kappaB: critical mechanisms in immune function and cancer.* Immunol Rev, 2012. **246**(1): p. 327-45.

[7] Pozhitkov, A.E., et al., *Tracing the dynamics of gene transcripts after organismal death.* Open Biol, 2017. **7**(1).

[8] Mandal, R., et al., *Caspase-8: The double-edged sword.* Biochim Biophys Acta Rev Cancer, 2020. **1873**(2): p. 188357.

[9] Fritsch, M., et al., *Caspase-8 is the molecular switch for apoptosis, necroptosis and pyroptosis.* Nature, 2019. **575**(7784): p. 683-687.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 505-511
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Identifying the genetic constraint characteristics and biological correlates of sudden unexplained death susceptibility genes

Junyi Lin[1], Chenyang Xu[1], Qi Shen[1], Ruiyang Tao[2], Shouyu Wang[1*]

[1]*Department of Forensic Medicine, School of Basic Medical Sciences, Fudan University, Shanghai, China.*

[2]*Shanghai Key Laboratory of Forensic Medicine, Shanghai Forensic Service Platform, Academy of Forensic Sciences, Ministry of Justice, Shanghai, China.*

** Corresponding author at: 131 Dong'an Road, 200032, Shanghai, China*

E-mail address: shouyu_wang@fudan.edu.cn (S. Wang)

Abstract

Sudden unexplained death (SUD) refers to sudden death in which the cause of death could not be established after a comprehensive medico-legal investigation, including pathological/toxicological assessment and forensic investigation of the circumstances of death. In the past decade, molecular autopsy has proved to be an efficient diagnostic tool in the multidisciplinary management of SUD. Though numerous pathogenic/likely pathogenic (P/LP) variants have been identified, there is still a large proportion of SUD cases without clear molecular autopsy findings, suggesting that the genetic predisposition contributing to SUD might be more complicated than expected. In this study, we analyzed published datasets to characterize the distribution pattern of rare P/LP variants. Enrichment analyses were subsequently conducted to identify the most affected canonical pathways/biological processes and phenotypes/cell types with strongest correlation to SUD. In addition, the constraint metrics of genes harboring P/LP variants were also assessed to investigate the role of selective pressure in shaping the genetic features of SUD.

Keywords

Sudden unexplained death (SUD), molecular autopsy, genetic constraint, pathogenic variants.

Introduction

Sudden unexplained death (SUD) constitutes a considerable portion of unexpected sudden natural death in children and young adults. Though primary arrhythmia

syndromes have been proposed as the most important risk factor for SUD, its pathophysiology and genetic background need to be further elucidated. Molecular autopsy has proved to be an efficient diagnostic tool in disclosing the genetic background of SUD. By post-mortem genetic testing, pathogenic/likely pathogenic (P/LP) variants in cardiovascular and metabolic genes were shown to exist in about one-third of the SUD cases. Based on published datasets, we aim to identify the genetic constraint characteristics and biological correlates of SUD susceptibility genes.

Material studied, methods, techniques

Our analysis included 12 studies focusing on the genetics of SUD, sudden arrhythmic death syndrome (SADS), or sudden infant death syndrome (SIDS) [1-12]. Enrichment analyses were conducted using Metascape [13] to identify the most affected canonical pathways/biological processes and correlated phenotypes/cell types. In addition, a comprehensive assessment of the constraint metrics and intolerance scores of genes harboring P/LP variants was performed to investigate the role of selective pressure in shaping the genetic features of SUD. Specifically, the Loss-of-function Observed/Expected Upper bound Fraction (LOEUF) [14], Residual Variation Intolerance Score (RVIS) percentile [15], protein-coding and non-coding Genomic Evolutionary Rate Profiling (pcGERP/ncGERP) percentile [16] was respectively evaluated. Continuous variables were presented as mean ± standard error of the mean (SEM). Comparisons between two groups were made using the Mann-Whitney U test. Pairwise multiple comparisons were made using the one-way ANOVA test and Dunn's multiple comparison post-test. All statistical tests were two-sided and a P value < 0.05 was considered statistically significant.

Results and discussion

In brief, 251 rare P/LP variants distributed in 113 different genes were identified from a total of 1272 cases. As shown in Figure 1A, RYR2 was found to be the most frequently affected gene (37 P/LP variants), followed by SCN5A (12 P/LP variants), TTN (12 P/LP variants), KCNH2 (10 P/LP variants) and KCNQ1 (8 P/LP variants). Previous studies already concluded that dysfunction of the critical calcium/sodium/potassium channel encoding genes could lead to ion channelopathies

manifested as severe inherited cardiac arrhythmias, including long QT syndrome, short QT syndrome, and Brugada syndrome [17]. In addition, the giant muscle filament titin encoding gene, TTN, was reported to be associated with different subtypes of cardiomyopathies [18].

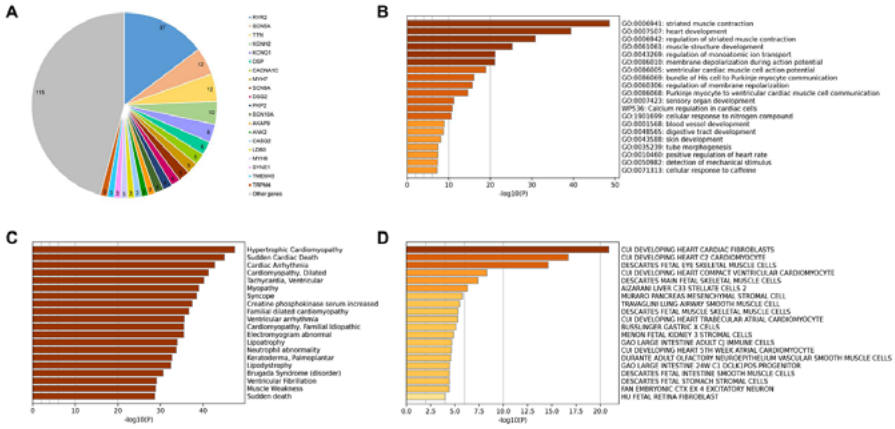


Figure 1. Gene-based distribution of P/LP variants and enrichment analysis results.

Among enriched pathways and processes, striated muscle contraction (GO: 0006941) was identified as the most relevant biological process (**Figure 1B**), followed by heart development (GO: 0007507) and regulation of striated muscle contraction (GO: 0006942). In terms of gene-disease association, hypertrophic cardiomyopathy was identified as the most common disease phenotype, followed by sudden cardiac death and cardiac arrhythmia (**Figure 1C**). These findings further confirmed that cardiac dysfunction is the most important pathologic basis of SUD. In addition, an enrichment analysis in Cell Type Signatures revealed that a majority of the 113 candidate genes were abundantly expressed in cardiac fibroblasts, cardiomyocytes, and fetal eye skeletal muscle cells (**Figure 1D**). These findings suggest that except for cell types known to be involved in cardiac dysfunction, disorders of other muscular organs might also be observed. Therefore, during the investigation of suspicious SUD, a thorough review of the medical history is recommended to obtain suggestive evidences supportive of molecular autopsy findings.

Regarding constraint metrics and intolerance scores, the 113 candidate genes harboring P/LP variants have consistently lower LOEUF, RVIS percentile, pcGERP percentile, and ncGERP percentile values compared with 113 random genes

from GENCODE Human Release 31 (**Figure 2**). Moreover, a stepwise comparison among genes that harbor different amount of variants (Group A: 1 variant; Group B: 2-4 variants; Group C: more than 5 variants) showed genes harboring more P/LP variants have lower LOEUF, RVIS percentile, pcGERP percentile, and ncGERP percentile values (**Figure 3**). Low LOEUF scores indicate strong selection against predicted loss-of-function (pLoF) variants in a given gene, while high LOEUF scores suggest a relatively higher tolerance to inactivation. Similarly, genes with lower RVIS percentile values are more intolerant to sequence variation compared to genes with higher RVIS percentile values. In addition, lower pcGERP percentile and ncGERP percentile values indicate that the genes have relatively strong dosage sensitivities due to their highly conserved protein-coding and non-coding sequences compared to the rest of the genome. Taken together, our findings suggest that genes reported to harbor substantial P/LP variants are relatively intolerant to deleterious genetic variants in either coding or non-coding region. Due to selective pressure, variants with functional effects in these genes have higher risks of leading to severe phenotypes before reproductive age. Correspondingly, the genetic constraint characteristics could be used as an important index for the interpretation of molecular autopsy findings, especially when there is a lack of evidence regarding the genotype-phenotype correlations.

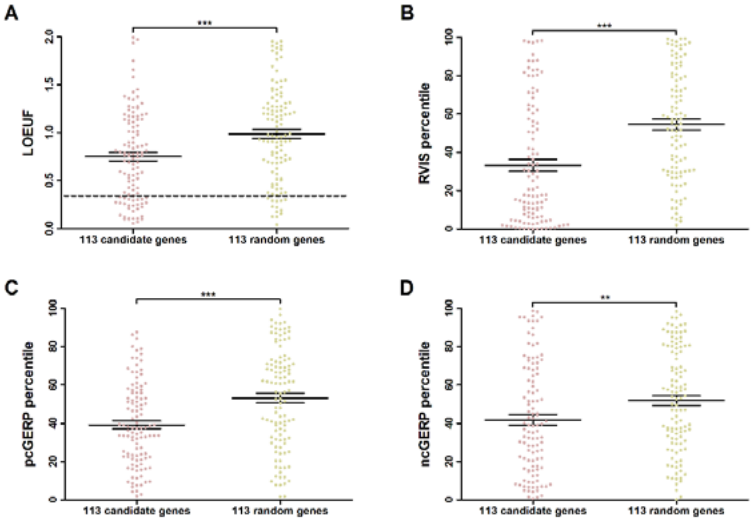


Figure 2. The 113 candidate genes have lower LOEUF, RVIS percentile, pcGERP and ncGERP percentile values compared with 113 random genes (**P ≤ 0.01; ***P ≤ 0.001).

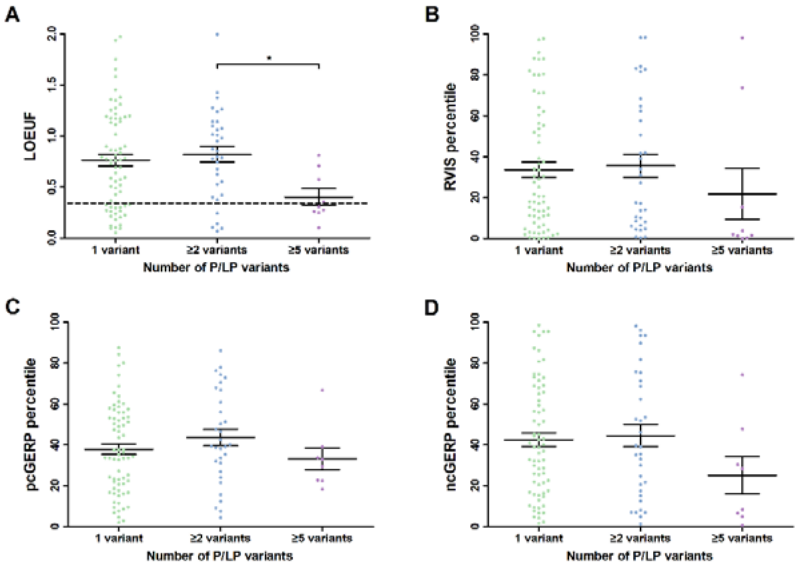


Figure 3. Genes harboring more P/LP variants have relatively lower LOEUF, RVIS percentile, pcGERP and ncGERP percentile values (*P≤0.05).

Conclusion

By analyzing published datasets from SUD/SADS/SIDS studies where a molecular autopsy was performed, the distribution pattern of 251 rare P/LP variants was characterized. Within the 113 genes harboring P/LP variants, ion channel encoding genes were found to be the most frequently affected genes. Among enriched pathways and processes, striated muscle contraction was identified as the most relevant biological process, while hypertrophic cardiomyopathy was identified as the most common disease phenotype. In addition to cardiac fibroblasts and cardiomyocytes, a majority of the candidate genes were also abundantly expressed in other cell types. More importantly, genes harboring more P/LP variants generally have lower LOEUF, RVIS percentile, pcGERP and ncGERP percentile values, suggesting that the genetic constraint characteristics might be used as an index when evaluating molecular autopsy results. In conclusion, our study provides insights into a better way of interpreting molecular autopsy findings.

Acknowledgments

This study was supported by the National Science Foundation of China (82101971) and the Opening Project of Shanghai Key Laboratory of Crime Scene Evidence (2018XCWZK21).

Conflict of interest statement

The authors declare that they have no conflicts of interest.

References

[1] J. Neubauer, M.R. Lecca, G. Russo, et al., Post-mortem whole-exome analysis in a large sudden infant death syndrome cohort with a focus on cardiovascular and metabolic genetic diseases, Eur J Hum Genet 25 (2017) 404-409.

[2] N. Lahrouchi, H. Raju, E.M. Lodder, et al., Utility of Post-Mortem Genetic Testing in Cases of Sudden Arrhythmic Death Syndrome, J Am Coll Cardiol 69 (2017) 2134-2145.

[3] D.J. Tester, L.C.H. Wong, P. Chanana, et al., Cardiac Genetic Predisposition in Sudden Infant Death Syndrome, J Am Coll Cardiol 71 (2018) 1217-1227.

[4] E.L. Salfati, E.G. Spencer, S.E. Topol, et al., Re-analysis of whole-exome sequencing data uncovers novel diagnostic variants and improves molecular diagnostic yields for sudden death and idiopathic diseases, Genome Med 11 (2019) 83.

[5] M.K. Larsen, S.L. Christiansen, C.L. Hertz, et al., Targeted molecular genetic testing in young sudden cardiac death victims from Western Denmark, Int J Legal Med 134 (2020) 111-121.

[6] N. Lahrouchi, H. Raju, E.M. Lodder, et al., The yield of postmortem genetic testing in sudden death cases with structural findings at autopsy, Eur J Hum Genet 28 (2020) 17-22.

[7] J. Neubauer, S. Wang, G. Russo, et al., Re-evaluation of single nucleotide variants and identification of structural variants in a cohort of 45 sudden unexplained death cases, Int J Legal Med 135 (2021) 1341-1349.

[8] M. Halvorsen, L. Gould, X. Wang, et al., De novo mutations in childhood cases of sudden unexplained death that disrupt intracellular Ca2+ regulation, Proc Natl Acad Sci U S A 118 (2021) e2115140118.

[9] U. Schön, A. Holzer, A. Laner, et al., HPO-driven virtual gene panel: a new efficient approach in molecular autopsy of sudden unexplained death, BMC Med Genomics 14 (2021) 94.

[10] S. Scheiper-Welling, M. Tabunscik, T.E. Gross, et al., Variant interpretation in molecular autopsy: a useful dilemma, Int J Legal Med 136 (2022) 475-482.

[11] S. Wang, Y. Chen, J. Du, et al., Post-mortem genetic analysis of sudden unexplained death in a young cohort: a whole-exome sequencing study, Int J Legal Med 137 (2023) 1661-1670.

[12] P. Votýpka, A. Krebsová, P. Norambuena-Poustková, et al., Post-mortem genetic testing in sudden cardiac death and genetic screening of relatives at risk: lessons learned from a Czech pilot multidisciplinary study, Int J Legal Med 137 (2023) 1787-1801.

[13] Y. Zhou, B. Zhou, L. Pache, et al., Metascape provides a biologist-oriented resource for the analysis of systems-level datasets, Nat Commun 10 (2019) 1523.

[14] K.J. Karczewski, L.C. Francioli, G. Tiao, et al., The mutational constraint spectrum quantified from variation in 141,456 humans, Nature 581 (2020) 434-443.

[15] S. Petrovski, Q. Wang, E.L. Heinzen, et al., Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes, PLoS Genet 9 (2013) e1003709.

[16] S. Petrovski, A.B. Gussow, Q. Wang, et al., The Intolerance of Regulatory Sequence to Genetic Variation Predicts Gene Dosage Sensitivity, PLoS Genet 11 (2015) e1005492.

[17] P.J. Schwartz, M.J. Ackerman, C. Antzelevitch, et al., Inherited cardiac arrhythmias, Nat Rev Dis Primers 6 (2020) 58.

[18] A.G. Jolfayi, E. Kohansal, S. Ghasemi, et al., Exploring TTN variants as genetic insights into cardiomyopathy pathogenesis and potential emerging clues to molecular mechanisms in cardiomyopathies, Sci Rep 14 (2024) 5313.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 513-519
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Unusual phenomena of 12058 individuals DNA profiling using commercial "A" reagent kit

Pol.Lt.Col.Dr. Salisa Prommakrit[1], Pol.Lt.Col.Dr. Niramon Masuntisuk[2] Pol.Capt. Tayawee Romgaew[2], Pol.Lt.Col. Sorada Pitilertpanya[2*]

[1]*Faculty of Forensic Science, Royal Police Cadet Academy, Nakhon Pathom, Thailand*

[2]*Central Police Forensic Science, Office of Police Forensic Science, Royal Thai Police, Bangkok, Thailand*

** Corresponding author at: Biology and DNA laboratory, Central Police Forensic Science, Office of Police Forensic Science, 14-15 building Royal Thai Police, Henry Dunant road, Wangmai, Pathumwan, Bangkok, Thailand, 10330 E-mail soradaka@gmail.com*

Abstract

Short Tandem Repeat (STR) in DNA comparison testing is used to link evidence to evidence or evidence to a person. The current testing loci, according to CODIS 2017, are 20 loci. For human DNA testing, the laboratory performs the commercial "A" reagent kit that has 24 testing loci: Amelogenin X, Amelogenin Y, DYS391, D1S1656, D2S441, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, D22S1045, CSF1PO, FGA, SE33, TH01, TPOX, and vWA. Those testing loci are stored in the DNA database system. Unusual phenomena, such as the occurrence of additional alleles and the disappearance of alleles, can occur in each DNA test. Therefore, the researcher is interested in the unusual phenomena using the commercial "A" test kit, and the purpose of this research is to determine the number of unusual phenomena occurring with the commercial "A" reagent kit. The research was conducted by testing the DNA of 12,058 persons during databasing by the Biology and DNA Subdivision, a laboratory of the Central Police Forensic Science Division under the Royal Thai Police. The test results found that there were two unusual phenomena. The first phenomenon detected in 114 DNA profiles (80 Thai and 34 Myanmar, 0.95%), was that one or two alleles had disappeared, but there was mystifying appearance of an additional allele in the adjacent locus DYS391. The other phenomenon detected in 12 DNA profiles (11 Thai and 1 Myanmar, 0.10%), was that the allele of locus DYS391 was missing, but an additional allele was displayed at locus TPOX. These phenomena may occur due to the lack of a supporting bin. The research findings show that it is necessary to be precautious when interpreting DNA using commercial kits.

Keywords
Autosomal Short Tandem Repeat, DNA profile, Allele, Forensic Genetics.

Introduction

The Biology and DNA Subdivision laboratory investigates biological stains and DNA on evidence, collecting STR DNA profiles using 24-loci commercial "A" test kit. These profiles are entered into a DNA database. Alleles of each locus were identified by converting DNA product sizes to STR alleles using genotyping allele bins.[1]. Unusual peaks, such as off-ladder peaks, arise from the absence of allele bins in the genotyping process, potentially resulting from the presence of variant alleles or extremely small alleles [1-3]. Unusual phenomena in the collected profiles, such as the occurrence of additional alleles or the absence of certain alleles, were observed during the analysis with the 24-loci commercial "A" test kit.

Therefore, the aim of this research is to determine the occurrence of unusual phenomena in the collected profiles using the 24-loci commercial "A" kit.

Material studied, methods, techniques

Collection of STR DNA Profiles

A total of 12,058 buccal swab samples were collected from consenting Thai and other national individuals.

Direct PCR

Lysis buffer and PCR reagents were used from the commercial "A" kit. For samples exhibiting unusual phenomena, analyses were repeated using commercial kit B or commercial kit C.

DNA Analysis

The DNA fragments were analyzed using the Genetic Analyzer 3500XL, and DNA profiles were generated using the program GeneMapper-IDX v.1.4.

Results

Among the 12,058 STR DNA profiles analyzed, we identified 126 profiles exhibiting two main unusual phenomena:

Disappearing Allele at D1S1656 and Appearance at DYS391:

We observed that the allele of locus D1S1656 had disappeared, but it was shown in locus DYS391 (Figure 1) in 114 STR DNA profiles (0.95% of 12,058 profiles, including 80 Thai and 34 Myanmar individuals).

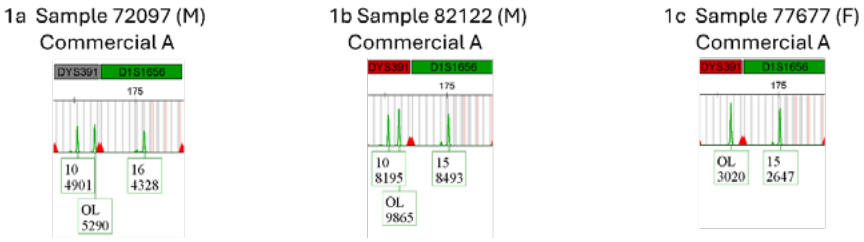


Figure 1. Shows alleles of both loci D1S1656 and DYS391 in the first phenomenon using Commercial A kit. An additional OL peak of DYS391 in DNA results male DNA sources, Sample 72097 (Figure 1a) and Sample 82122 (Figure 1b), and a female DNA source Sample 77677 (1c); causing the first phenomenon.

We re-amplified the aforementioned samples using commercial B kit and detected the presence of additional alleles 8 and 9 at the D1S1656 locus (Figure 2).

Figure 2. Shows alleles of locus D1S1656 in DNA results of Sample 72097 and Sample 82122 using Commercial B kit. A well-balanced heterozygous peak of locus D1S1656 of both Sample 72097 (Figure 2a) and Sample 82122 (Figure 2b).

Missing Allele at DYS391 and Appearance at TPOX:

We found that the allele of locus DYS391 was missing while it was displayed as an additional allele at locus TPOX (Figure 3) in 12 STR DNA profiles including 11 Thai and 1 Myanmar (0.10% of 12,058 profiles).

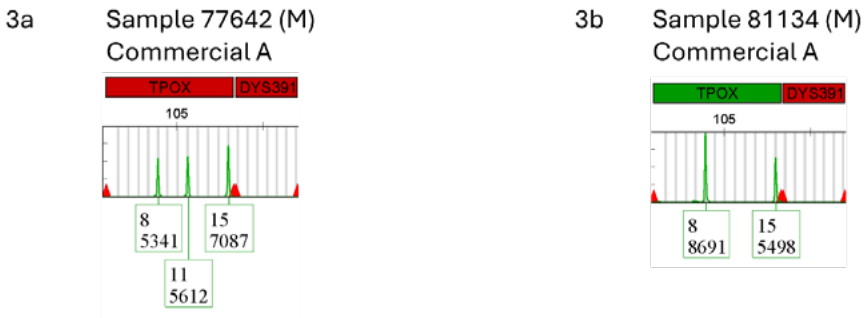


Figure 3. Shows alleles of both loci DYS391 and TPOX in the second phenomenon using Commercial A kit. Figure 3a shows a tri-allele occurrence at locus TPOX of Sample 77642 (male). Figure 3b shows a pseudo-heterozygous peak at locus TPOX of Sample 81134 (male), while the allele of DYS391 was missing.

A total of 12 STR DNA profiles were re-amplified using commercial C kit. The results revealed the absence of allele 15 at the TPOX locus, while allele 8 was identified at the DYS391 locus (Figure 4).

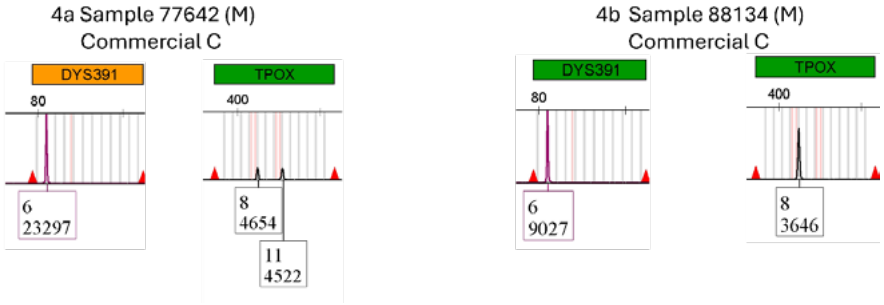


Figure 4. Shows alleles of both loci DYS391 and TPOX in the second phenomenon using Commercial A kit. Figure 3a shows a tri-allele occurrence at locus TPOX of Sample 77642 (male). Figure 3b shows a pseudo-heterozygous peak at locus TPOX of Sample 81134 (male), while the allele of DYS391 was missing.

Discussion

The first phenomenon may occur due to the lack of a genotyping allele bins for D1S1656, as 8 and 9 are found to be relatively rare (Table 1) [3-6], causing

displacement of an allele to an adjacent locus DYS391 in the same panel. Similarly, a missing allele from locus DYS391 of a male DNA source observed in the second phenomenon was shifted to locus TPOX causing an unusual tri-allele or pseudo-heterozygous peak. Both phenomena are "inter-loci shift" that may lead to misinterpretation of DNA result (Figure 5). An inter-loci shift was also identified as being associated with an extremely short allele of D1S1656 utilized in the PowerPlex® 16 ESI and ESX kits [3].

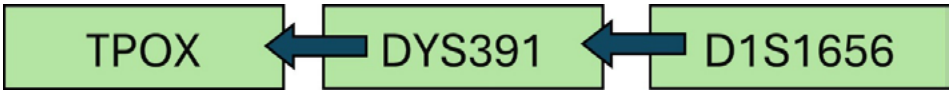


Figure 5. shows "inter-loci shift" phenomena between loci TPOX, DYS391 and D1S1656 using Commercial A; Arrows in figure 5a represent the direction of "inter-loci shift" found in this research.

Table 1. *Allele frequencies of rare alleles [4-7]*

Locus	Allele Frequency			
D1S1656	Thailand[a]		Europe[b]	
	8	9	8	9
	0.0042	0.0012	No data	0.0000707
DYS391	Thailand[c]		YHRD database[d]	
	6		6	
	0.0049		0.0023	

a. Autosomal STR allele frequencies from Thai population [4]
b. Autosomal STR allele frequencies from European populations [5]
c. Y STR allele frequencies from Thai population [6]
d. Y STR allele frequencies from YHRD database [7]

Conclusion

We identified unusual phenomena in 126 of 12,058 profiles from Thai and Myanmar individuals using the commercial "A" kit (approximately 1%). Therefore, it is necessary to be cautious when interpreting DNA results using the kit with DNA profiles in Thailand and Myanmar population, especially when interpreting DNA results of an unknown DNA sources lacking information on gender.

Acknowledgments

We would like to express our gratitude to the Biology and DNA Subdivision of the Central Police Forensic Science, Office of Police Forensic Science, Royal Thai Police, for providing the laboratory facilities. Special thanks are extended to Pol. Capt. Natcha Sanpaeng for their invaluable assistance and unwavering support during all challenging situations.

Conflict of interest statement

The authors declare that there are no conflicts of interest regarding the publication of this article. All financial, personal, and professional relationships that could be perceived as influencing the research outcomes have been disclosed.

If any potential conflicts arise during the course of the study or its publication, they will be reported and managed in accordance with the ethical standards of the Proceedings of the 30th Congress of the International Society of Forensic Genetics.

References

[1] J.M. Butler, *Advanced topics in forensic DNA typing: Methodology*. Elsevier Academic Press, 2012.

[2] M.C. Kline, C.R. Hill, A.E. Decker, J.M. Butler, STR sequence analysis for characterizing normal, variant, and null alleles. *Forensic Science International: Genetics* 5, (2011) 329-332. https://doi.org/10.1016/j.fsigen.2010.09.005

[3] C. Oz, A. Karageorgos, M. Kocak, Recognizing the advantage of employing the PowerPlex® ESI and PowerPlex® ESX complementary kits for purposes of clarification in routine casework. *Forensic Science International: Genetics* 7, (2013) 461–466. https://doi.org/10.1016/j.fsigen.2013.02.001

[4] N. Boonderm, D. Suriyanratakorn, S. Sangpueng, N. Onthong, A. Nettakul, W. Waiyawuth, Population genetic data of 21 STR markers in Thais of southern border provinces of Thailand. *Forensic Science International: Genetics Supplement Series* 6, (2017) e523–e525. https://doi.org/10.1016/j.fsigss.2017.09.205

[5] M. Bodner, W. Parson, The STRidER report on two years of quality control of autosomal STR population datasets. Genes 11, (2020) 901. https://doi.org/10.3390/genes11080901

[6] T. Siriboonpiputtana, U. Jomsawat, T. Rinthachai, J. Thanakitgosate, J. Shotivaranon, N. Limsuwanachot, P. Polyorat, B. Rerkamnuaychoke, Y-chromosomal STR haplotypes in Central Thai population. *Forensic Science International: Genetics* 4, (2010) e71–e72. https://doi.org/10.1016/j.fsigen.2009.06.007

[7] S. Willuweit, L. Roewer, Y chromosome haplotype reference database (YHRD): Update. *Forensic Science International: Genetics* 1, (2007) 83-87. https://doi.org/10.1016/j.fsigen.2007.01.017

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 521-527
DOI: https://dx.doi.org/10.15304/cc.2025.1869

African genetic and linguistic origins of Palenque, Colombia

Martínez Beatriz[1*], Marrugo Javier[1], Jordan I. King[2,3]

[1]*Instituto de Investigaciones Inmunológicas, Universidad de Cartagena, Cartagena, Colombia.*
[2]*IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, Georgia, USA.*
[3]*School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA.*

** Corresponding author at: Beatriz Martinez. Instituto de Investigaciones Inmunológicas, Universidad de Cartagena, Cartagena, Colombia. Cell phone number: +57 3043467817*
Email: bmartineza1@unicartagena.edu.co

Abstract

The Colombian town of San Basilio de Palenque, established as a refuge for enslaved Africans more than 400 years ago, has maintained an African culture and traditions that include the Afro-Spanish Creole language. It was previously discovered that the Creole language of the Palenque population has a single African substrate: the Yombe language of the Pointe-Noire region in the Republic of the Congo (Ansari-Pour, N. et al., 2016). We characterized the African subcontinental origins of Palenque individuals, through fine-scale analysis of WGS data from Palenque donors along with African reference samples using Chromopainter v2. While individuals from Palenque show the highest levels of Bantu (Loango Coast) ancestry, it is also clear that they have substantial contributions from the three African regions that participated in the transatlantic slave trade to Colombia (51% Bantu, Loango Coast, ancestry), 35% of Ensenada Benin ancestry and 14% of Senegambia ancestry). Palenque also shows substantial genetic diversity within the Bantu genetic ancestry component. The Yombe language corresponds to the Kongo-like population group that constitutes 25% of the Palenque ancestry, while the remaining Bantu population groups represent 26% of the Palenque ancestry. With our results, we could conclude that the African substrate of *Palenquero* language is actually more complex than previously imagined, with contributions from multiple African languages.

Keywords

Afro descendant population, creole language, WGS, ancestry.

Introduction

Widespread *cimarronaje* during the colonial period led to the formation of autonomous communities of enslaved Africans and their descendants (maroons) throughout the Americas [1]. Maroon communities took refuge in fortified villages known as *palenques*, which were strategically located in inhospitable and hard to reach regions during the 16th and 17th centuries [2]. The town of San Basilio de Palenque is the only such community that exists to this day, more than 400 years after it was established [2-3]. Palenque is located in the Montes de Maria region, just over 40 miles southeast of the port city of Cartagena in Colombia's Bolivar department. To this day, the residents of Palenque speak a creole language known as *Palenquero*, which combines Spanish and African linguistic elements [4,5]. Linguists have characterized a single African substrate for *Palenquero* – the Kikongo language, which is a member of the Bantu language family spoken in the Republic of the Congo, the Democratic Republic of Congo, and Angola [4-6] The Yombe language, from the Pointe-Noire area of the Republic of the Congo, is considered to be the specific source of *Palenquero* [7,8]. Previous studies on the genetic ancestry of Palenque have a limited resolution on the African ancestry of Palenque [9, 10-15]. For this study, we performed a whole genome sequence (WGS) analysis to more deeply understand the African genetic origins of Palenque and considered genetic ancestry in the context of African linguistic groups.

Material studied, methods, techniques.

Material I

DNA sample donors were enrolled by the Institute of Immunological Research of the University of Cartagena. All donors provided written informed consent. Whole genome sequencing of 34 Palenque samples to 30x coverage was performed on the Illumina HiSeq 2000, as part of the Consortium on Asthma among African Ancestry Populations in the Americas (CAAPA) project (https://www.caapa-project.org/), and variants were called as previously described [16].

Material II

The program Chromopainter version 2 was used to infer subcontinental ancestry admixed American genomes using ancestry-specific haplotypes compared to ancestry group-specific reference samples [17]. Chromopainter was run on RFMix

masked genomic data to characterize each continental ancestry component of admixed genomes – African, European, and Native American – separately.

Results

We characterized subcontinental African genetic ancestry in the context of the geographic distribution of African language groups making the fine-scale genetic ancestry of 35 African populations coming from the three main regions that participated in the transatlantic slave trade to Colombia: Senegambia (West Africa), the Bight of Benin (West Central Africa), and the Loango Coast (Southwest Africa) (Figure 1A). Additional Rain Forest Hunter Gatherer (RFHG) and East Bantu populations were included for comparison. The major axis of genetic diversity revealed by PCA divides the Senegambia and the Bight of Benin population groups from the Loango Coast group, which corresponds to the distinction between non-Bantu and Bantu African language groups, respectively (Figure 1B). The results of Chromopainter analysis show largely coherent patterns of genetic ancestry within the non-Bantu populations from Senegambia and the Bight of Benin; whereas the Loango Coast populations show substantial population structure with five distinct Bantu sub-groups (Figure 1C). Palenque show 51% of Bantu (Loango Coast) ancestry, 35% of Ensenada Benin ancestry and 14% of Senegambia ancestry (Figure 1D). Palenque also shows substantial genetic diversity within the Bantu genetic ancestry component (Figure 1E). The Yombe language corresponds to the Kongo-like population group analyzed here, which makes up 25% of Palenque ancestry, while the remaining Bantu population groups make up 26% of Palenque ancestry.

Discussion

For this study, we were most interested to ask about the sub-continental African origins of *Palenqueros*. We hypothesized that Palenque should show a mix of genetic ancestry from different regions of Africa with high levels of genetic diversity. Our results show that Palenque have a majority of Bantu African ancestry, they also have substantial ancestry contributions from Senegambia and the Bight of Benin. Furthermore, within the Bantu ancestry component, Palenque shows evidence of contributions from numerous groups beyond the single Kongo-like group that corresponds to the source of the African language substrate. Altogether, the

population of Palenque shows ancestral affinities to African populations that speak a wide variety of languages from the diverse Niger-Congo language family. These results highlight a striking dissonance between the African genetic origins of Palenque and the cultural dominance of a single African language substrate for the creole Palenquero language. The Kikongo language that forms the African substrate for Palenquero was spoken in the African kingdom of Kongo, which dominated much of southwestern Africa from the late 14th to the early 20th centuries [18]. By far, the largest number of enslaved Africans, close to 5.9 million, were taken to the Americas from this region (Loango Coast) during the transatlantic slave trade (https://www.slavevoyages.org/). Kikongo was, and remains to this day, the main base of a number of creole dialects spoken in adjacent African regions [19]. Once established as the dominant African language substrate by the founders of Palenque, later arrivals may have been compelled to adopt Palenquero, including both its African and Spanish strata, irrespective of their native African tongue.

Conclusion

With our results, we could conclude that the African substrate of Palenquero is actually more complex than previously imagined, with contributions from multiple African languages. Nevertheless, our results could also be used to motivate further studies and potentially new discoveries about the African substrate of Palenque.

Acknowledgments

We thank all sample donors for their contribution to this work, and all the people who helped with samples collection, namely Regina Miranda from the Community of San Basilio de Palenque, the personal of the Benkos Biojó Rural School, and Miguel Obeso from the *etno educación* program. The University of Cartagena, Colombia, for the Consortium on Asthma among African Ancestry Populations in the Americas (CAAPA) project, USA, and IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, Georgia, USA supported this work.

Conflict of interest statement

The authors declare no competing interests.

References

[1] Price, R. *Maroon societies: Rebel slave communities in the Americas*, Johns Hopkins University Press (Eds), Baltimore, MD, 1996

[2] Navarrete, M.C., Palenques: Maroons and Castas in Colombia's Caribbean Regions, in: Schwegler, A., Kirschen, B. & Maglia, G (Eds), *Orality, Identity, and Resistance in Palenque (Colombia): An interdisciplinary approach,* Amsterdam, Netherlands, 2017 pp 265-293

[3] Friedemann, N. & Patiño Rosselli, C, *Lengua y sociedad en el Palenque de San Basilio*, Instituto Caro y Cuervo, (Eds), Bogotá, Colombia, 1983

[4] Escalante, A, *El Palenque de San Basilio: una comunidad de descendientes de negros cimarrones*, Mejoras, Barranquilla (Eds), Colombia, 1979.

[5] Granda Gutiérrez, G.d., Sobre la procedencia africana del habla'criolla'de San Basilio de Palenque (Bolívar, Colombia). *Thesaurus: Boletín del Instituto Caro y Cuervo* 26, 84-94,1971.

[6] Schwegler, A, Palenque (ro): the search for its African substrate, in: Lefebvre, C *Creoles (Eds), their substrates and language typology,* Amsterdam, Netherlands, 2011, pp 225-249

[7] Schwegler, A, Combining population genetics (DNA) with historical linguistics, in: Sessarego, S. & Tejedo-Herrero, F (Eds), *Spanish language and sociolinguistic analysis*, Amsterdam, Netherlands, 2016 pp 33-88

[8] Schwegler, A, On the African origin (s) of Palenquero, in: Schwegler, A., Kirschen, B. & Maglia, G (Eds), *Orality, Identity, and Resistance in Palenque (Colombia): An interdisciplinary approach,* Amsterdam, Netherlands, 2017). pp 51-119

[9] Ansari-Pour, N. et al. Palenque de San Basilio in Colombia: genetic data support an oral history of a paternal ancestry in Congo. Proc Biol Sci 283 (2016) 20152980.

[10] Martinez, B. *et al.* Searching for the roots of the first free African American community. *Sci Rep* 10 (2020) 206-34

[11] Arnaiz-Villena, A. *et al.* The first free Africans in America: HLA study in San Basilio de Palenque (Colombia). *Hum Immunol* 79 (2018) 585-586

[12] Martínez, B. *et al.* Population genetic data of 38 autosomal InDels in San Basilio de Palenque, the first free town in America. *Forensic Science International: Genetics Supplement Series* 4 (2013) e73-e74

[13] Martínez, B. *et al.* Ancestry estimates in afrodescendant population from San Basilio de Palenque, Colombia. *Forensic Science International: Genetics Supplement Series* 6 (2017) e224-e225

[14] Noguera, M.C. *et al.* Colombia's racial crucible: Y chromosome evidence from six admixed communities in the Department of Bolivar. *Ann Hum Biol* 41, (2014) 453-9

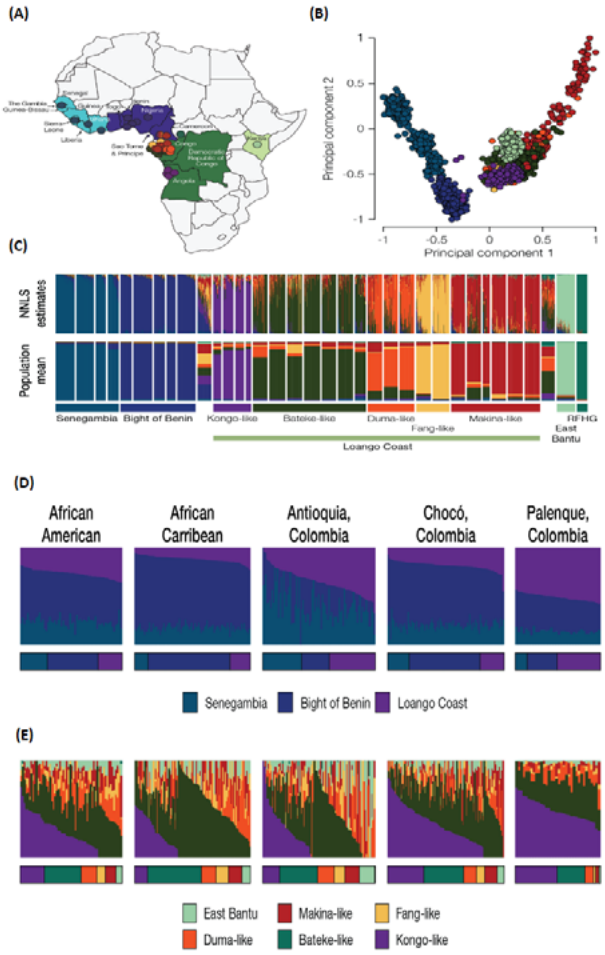
[15] Solé-Llussà, A. *et al.* Exploring the African genetic influence in the first settlement founded by African slaves in America. *Anthropological Science* 123 (2015)121-128

[16] Mathias, R.A. *et al.* A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. *Nat Commun* 7 (2016) 12522

[17] Lawson, D.J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. PLoS Genet 8 (2012) e1002453

[18] Thornton, J.K. *Africa and Africans in the Making of the Atlantic World, 1400-1800*, Cambridge University Press (Eds), Cambridge, 1998

[19] Mufwene, S.S. Kikongo-Kituba, in: *Encyclopedia Britannica,* Encyclopedia Britannica, (Eds), Edinburgh, 2008



African genetic ancestry and language groups. (A) Locations of African reference populations among the three main sampling regions: West African (Senegambia, light blue), West Central Africa (Bight of Benin, dark blue), and Southwest Africa (Loango Coast, green). (B) PCA showing genetic relationships among African reference populations (colored as shown in panel A). (C) Chromopainter analysis showing genetic structure among African reference populations, grouped according to broad geographic regions and population groups. Individual level NNLS ancestry estimates are shown along with population average estimates. (D) Individual level and population average NNLS ancestry estimates for the three main African sampling regions: Senegambia (light blue), Bight of Benin (dark blue), and Loango Coast (purple). (E) Individual level and population average NNLS ancestry estimates for the Loango Coast (Bantu) component of African genetic ancestry.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 529-533
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Genetic Relationship between two sub-populations within the Akan ethnic group in Ghana

Sandra Appiah Kubi, William H. Goodwin

University of Central Lancashire, Preston, United Kingdom

Corresponding author at: SAppiah-kubi@uclan.ac.uk.

Abstract

Short tandem repeats (STRs) continue to be the golden standard used in forensic genetics. While many population studies have been published, data on sub-Saharan African populations is limited. In Ghana there are 75 recognized ethnic groups; the largest ethnic group is the Akan, who make up nearly 50% of the population.

This study involved sampling two populations of the Akan population, one from the Ashanti region and one from the Eastern region, with a total of 196 samples from unrelated males and females. These were extracted and analysed using the GlobalFiler™ Amplification kit. Data were analysed to test Hardy-Weinberg equilibrium (HWE) using Arlequin and population differentiation using STRUCTURE,

STRUCTURE analysis did not detect any significant differences between the two subpopulations and F_{ST} between the two populations was 0.00145, indicating a negligible level of genetic differentiation.

Keywords

STRs, GlobalFiler Amplification Kit, Ghana, Akan, Population genetics.

1. Introduction

Ghana is situated in west Africa bordered by Burkina Faso, Togo and Côte d'Ivoire, and on its southern boarder the Atlantic Ocean. The land area is approximately 238,000 km^2 and the population is estimated to be just over 32,000,000. Historical geographers and cultural anthropologists have categorised the Ghanaian population into five primary groups: Akan, Ewe, Mole-Dagbani, Guan, and Ga-Dangbe, based on cultural and linguistic criteria [1]. The principal ethnic groups in Ghana

comprise Akan at 45.7% of the population, Mole-Dagomba at 18.5%, Ewe at 12.8%, Ga-Dangme at 7.1%, Gurma at 5.7%, Guan at 3.2%, Grusi at 2.7%, Konkomba at 3.5%, and Kusaasi at 1.2%. Approximately 70 percent of the entire population resides in the southern region of the country [1].

Limited forensic studies have been undertaken in Ghana. Wepeba et al. [2] performed a population genetic analysis of the four principal ethnic groups in Ghana, utilising the 21 loci GlobalFiler PCR amplification kit.

The study sought to analyse two populations of the Akan demographic: one from the Ashanti region and another from the Eastern region, comprising a total of 196 samples from unrelated males and females (98 samples each) to assess genetic diversity within the Akan ethnic group in Ghana.

2. Material studied, methods, techniques

2.1. Sample collection, DNA extraction and Quantification

A total of 196 buccal swabs were collected from 2 Akan sub-populations with informed consent from participants. Ethical approval for the study was granted by the University of Central Lancashire's Ethics review panel and from the College of Health Sciences, School of Medicine and Dentistry and the Committee on Human Research, Publications and Ethics of Kwame Nkrumah University of Science and Technology (KNUST), Ghana. DNA was extracted using the Puregene® DNA Extraction Kit (Qiagen UK) following the manufacturer's protocol. The Qubit 3.0 fluorometer was used to quantify DNA using the high sensitivity dsDNA Qubit Kit (Thermofisher Scientific, USA). Agarose gel electrophoresis was used to evaluate the purity, and the molecular weight of the DNA extracted.

2.2.PCR Amplification and Fragment Analysis

PCR amplification was performed using the GlobalFiler™ amplification kit according to manufacturer's protocol with a reduced volume of 7.5 µl with 1 µl (0.5 ng) of template DNA. The Applied Biosystems 3500 Genetic Analyzer was used for capillary electrophoresis and profiles were analysed using GeneMapper I-DX v. 1. 4.

2.3. Data Analysis

Genetic distance, Hardy-Weinberg equilibrium (HWE) and p-values were tested using the Arlequin software v 3.5 [3] and STRUCTURE 2.0 software [4] was utilised to do the population cluster analysis.

3. Results

3.1 Hardy Weinberg Equilibrium

Arlequin software calculated the p-values for the 21 STR loci assessed in the two sub-populations. The loci that deviated from Hardy-Weinberg equilibrium are marked in red; however, no deviations were observed following the Bonferroni correction (0.002).

Table 1. Shows the p-values for the exact of the 21 STR loci tested in the two sub-populations generated using Arlequin software.

Locus	Akan (Ashanti)P-Values	Akan (Eastern)P-Values
CSF1PO	0.1746	0.3185
D10S1248	0.7961	0.226
D12S391	0.0792	0.5155
D13S317	0.4378	0.4519
D16S539	0.9617	0.4914
D18S51	0.0449	0.6862
D19S433	0.0428	0.0993
D1S1656	0.9875	0.0758
D21S11	0.4882	0.0049
D22S1045	0.2983	0.2725
D2S1338	0.5796	0.7951
D2S441	0.3844	0.3846
D3S1358	0.8335	0.9834
D5S818	0.2917	0.6771
D7S820	0.3651	0.8634
D8S1179	0.0747	0.156
FGA	0.7157	0.335
SE33	0.3762	0.8395
TH01	0.8556	0.5361
TPOX	0.2813	0.4395
vWA	0.383	0.4065

F_{ST} between the two Akan populations was calculated using Arlequin to be 0.00145.

3.2 Cluster Analysis

The study utilized Structure analysis software to evaluate population clusters for the studied subpopulation, a published Akan population [2] and Saudi Arabian population [5] determining K between 2 and 4 using the admixture and correlated allele frequencies model. A Saudi Arabian population was used as a non-African population in the analysis. The results (Figure 1) did not detect significant levels of sub-structure between the Akan populations

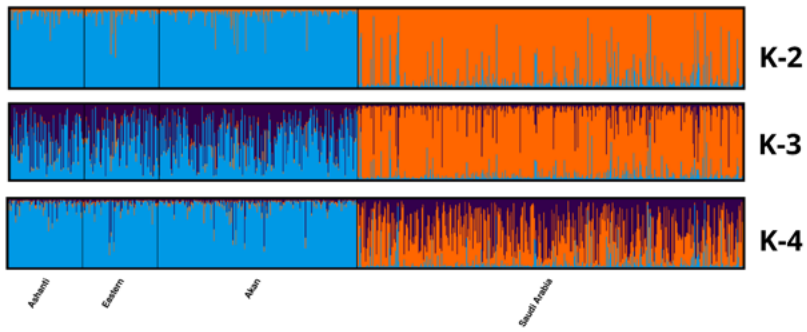


Figure 1. The CLUMPAK software has enhanced the Structure cluster image of the two studied populations, and a previously published Akan population. The distinct K clusters are shown for K-2 to K-4.

Discussion

The GlobalFiler™ has demonstrated suitability for forensic analysis in the examined Akan subpopulation, due to the statistical independence of the 21 loci in the GlobalFiler™ Amplification kit.

D18S51, D19S433, D21S11 of the 21 GlobalFiler PCR loci deviated from HWE, no significant deviation was seen following the use of the Bonferroni correction (0.002).

STRUCTURE analysis study revealed no substantial differences between the two Akan sub-populations studied here or another previously studies Akan population. Genetic pairwise F_{ST} value of 0.00145 between the Ashanti Akan and Eastern Akan populations within Ghana, indicated a negligible genetic differentiation. Akan's practice matrilineal system, however, in the present-day Ghana intermarriages are very common among all ethnic groups and subgroups including Akan Ashanti and Akan Eastern.

Conclusion

The evaluation of autosomal STRs in the Ghanaian population using the GlobalFiler PCR amplification kits specifically the two Akan sub-population for kinship and forensic testing has yielded robust results. This will serve as a reference data to use for forensic casework and general human identification in Ghana.

Acknowledgments

This project was fully funded by the Ghana National Petroleum Corporation Foundation.

Conflict of interest statement

The authors have no conflicts of interest

References

[1] Ghana Statistical Service. Ghana Population and Housing Census. Accra (2021) Available at: https://www.statsghana.gov.gh.

[2] P. Wepeba, A. Iyengar, and W. Goodwin, Heterozygous 21 STR loci and triplet alleles observed in population genetic analysis of the GlobalFiler STR loci in the Ghanaian population, Foren. Sci. Int.: Genet Suppl. Ser 7(1): (2019)753-754.

[3] L. Escoffier, HE. Lischer. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources10(3): (2010) 564-7.

[4] JK. Pritchard, M. Stephens, P. Donnelly. Inference of population structure using multilocus genotype data. Genetics. 155(2): (2000) 945-59

[5] HM. Alsafiah, WH. Goodwin, S, Hadi, MA. Alshaikhi PP. Wepeba. Population genetic data for 21 autosomal STR loci for the Saudi Arabian population using the GlobalFiler® PCR amplification kit. Forensic Science International: Genetics. 1;31: (2017) e59-61.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 535-540
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Analysis of the Ancestry and admixture proportions in Peruvian Mestizo Populations using an Insertion Deletion multiplex

Carlos Neyra-Rivera[1*], Rosmery Lopez Sam[2,3], Edgardo Delgado Ramos[4], Jhonny A. De La Cruz-Vargas[1], Leonor Gusmão[5]

[1]*Instituto de Investigaciones en Ciencias Biomedicas, Universidad Ricardo Palma, Lima, Peru.*
[2]*Posgrado Facultad de Ciencias Biologicas, Universidad Nacional Mayor de San Marcos, Lima, Peru.*
[3]*Gendx Latinoamerica, Lima, Peru.*
[4]*Instituto de Medicina Legal y Ciencias Forenses, Lima, Perú.*
[5]*DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil.*

** Corresponding author at: Av. Alfredo Benavides 5440, Santiago de Surco, Lima 15039, carlosdavidmp@outlook.es*

Abstract

In the present investigation, we aim at estimate African, European, Native American and Asian ancestry proportions in Peruvian mestizo populations, using 46 ancestry informative Insertion/Deletion polymorphisms (AIM-INDELs). A total of 172 unrelated mestizos from the 3 main geographic regions of Peru (Coastal, Amazon rainforest and Andes highlands) were analyzed and compared with data from 273 African, 346 European, 64 Native American, and 291 Asian individuals. All samples were genotyped for the 46 AIM-INDELs in a single PCR multiplex. Based on the results from admixture analyses, it was possible to observe a predominance in the Native Americans ancestry in the overall Peruvian mestizo population, as well as in each country region. A non-differentiation analyses based on pairwise FST genetic distances showed significant differences between the population samples from the costal area and the Andes highlands.

Keywords

Insertion–deletion polymorphisms, mestizo, Native American.

1. Introduction

Peru is a country with high population diversity. According to the census conducted in 2017 by the National Institute of Statistics and Informatics [1], a total

of 55 indigenous peoples were identified (51 from the Amazon and 4 from the Andes highlands). This population diversity has been investigated using autosomal and Y-chromosomal STR markers in mestizo, Aymara, and jungle populations of Peru [2-11]. Due to their high mutation rates, the STRs tend to present a high variation inside rather than within populations. Therefore, they are not the most suitable markers to estimate the different proportions of continental contribution in admixed populations. Genetic studies using other type of markers carried out in Peruvian population are still a scarce. Therefore, the objective of the present investigation was to estimate African, European, Native American, and Asian ancestry proportions in Peruvian mestizo populations, using a panel of 46 Insertion/Deletion polymorphisms (AIM-INDELs), which was design to estimate European, African, Native American, and East Asian ancestry proportions [12].

2. Material studied, methods, techniques

This study was approved by the ethics committee of the Institute of Tropical Medicine "Daniel Alcides Carrion" from the Universidad Nacional Mayor de San Marcos (Certificate of Approval CIEI-2018-015), and all participants freely and voluntarily signed an informed consent. In the present investigation, 172 unrelated mestizos from the 3 main geographic regions of Peru (36 from the Coastal area, 8 from the Amazon rainforest and 128 from the Andes highlands) were analyzed and compared with data from 273 African, 346 European, 64 Native American and 291 Asian individuals. A blood sample was taken by finger puncture from each donor and placed on Nucleid Cards (Copan). Then, a two 3.0mm punch was taken from each card and genomic DNA was extracted using 5% Chelex resin. The DNA concentrations in each sample was quantified using the Qubit. All samples were genotyped for 46 AIM-INDELs in a single PCR multiplex described by Pereira et al. [11], after adjusting the total reaction volume to 5 µL, and using 2 ng of template DNA. The following PCR thermocycling conditions were used: an initial step of 15 min at 95 ºC, followed by 28 cycles at 94 ºC for 30 s, 60 ºC for 90 s, 72 ºC for 45 s and a final extension at 72 ºC for 60 min. The amplified products were separated and detected in an Applied Biosystem ™ 3500XL Genetic Analyzer (Life Technologies). Genotyping was performed using the GeneMapper® ID-X v1.5 software. Ancestry proportions were estimated for each individual with the Structure Software [13], for k=4, and using reference population data from 273 African, 346 European, 64 Native American and 291 Asian individuals. The software Arlequin ver 3.5.2.2 [14] was used to

perform population pairwise comparisons using FST values, and to calculate the corresponding non-differentiation p-values.

3. Results and Discussion

The estimates of different continental contributions are indicated in Table 1. In the overall mestizo Peruvian population, the main contribution was from Native America, followed by European, and, to a lesser extent, African and East Asian contributions. The Native American ancestry predominate in the three studied regions, although being lower in the Coastal area. The European ancestry showed the second-highest proportion values in individuals from the Coastal area and Andes highlands. Meanwhile, Asian proportion presented the lowest values in the Coastal area and Andes highlands, and African proportion was the lowest in the Amazon rainforest. The Asian proportion was higher in the Amazon rainforest than in the other two regions (Table 1). Pairwise population comparisons between the three Peruvian population groups, and European, African, Native America, and East Asian populations showed significant FST values in all comparisons, except between the Amazon rainforest and the other two regions of Peru. It is important to note, however, that the Amazon sample only has 8 individuals, so this result should be analyzed with caution and new tests should be carried out on a larger sample (Table 2).

Table 1. Estimates of ancestry proportions in mestizo Peruvian population. The ancestry proportions of the mestizo population of Peru, from the Coastal area, Andes highland, and Amazon rainforest, were estimated considering reference populations from Africa (AFR), Europe (EUR), East Asia (EAS) and Native America (NAM).

SAMPLES	AFR	EUR	NAM	EAS
Costal area	0.078	0.369	0.484	0.069
Andes highlands	0.044	0.201	0.713	0.041
Amazon rainforest	0.042	0.190	0.663	0.106
Peru mestizo	0.051	0.236	0.663	0.050

Table 2. Pairwise comparisons between populations from the three geographical regions of Peru, and Europe, Africa, Native America, East Asia. The F_{ST} values are represented below the diagonal, and the corresponding *p*-values are presented above the diagonal.

	EUR	NAM	AFR	EAS	Costal area	Amazon rainforest	Andes highlands
Europe	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
Native America	0.2859	*	<5E-6	<5E-6	<5E-6	0.0011	<5E-6
Africa	0.3243	0.4197	*	<5E-6	<5E-6	<5E-6	<5E-6
East Asia	0.3035	0.2513	0.3977	*	<5E-6	<5E-6	<5E-6
Costal area	0.1193	0.0828	0.3318	0.2165	*	0.4399	<5E-6
Amazon rainforest	0.1852	0.0430	0.3706	0.2193	0.0012	*	0.8235
Andes highlands	0.1919	0.0343	0.3725	0.2257	0.0172	0.0071	*

Number of permutations: 50175; s.e. ≤ 0.0021

4. Conclusion

Based on the results, the proportion of Native Americans predominates in the Peruvian mestizo population, indicating that despite the migrations Peru has experienced (from Europe, Africa and Asia), it still maintains a high proportion of the Native American component. Furthermore, it was possible to observe differences between the 3 regions, with the Andes and coastal regions showing statistically significant differences.

5. Acknowledgments

This present research was supported by the Instituto de Investigaciones de Ciencias Biomédicas of the Universidad Ricardo Palma. L.G. is supported by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ; CNE-2022) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; ref. 315007/2023-0), Brazil.

6. Conflict of interest statement

None

7. References

[1] Instituto Nacional de Estadistica e Informatica, Perú: Estado de la Población en el año del Bicentenario, 2021, 2021 https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1803/libro.pdf

[2] E. Delgado, C.D. Neyra, Allele frequencies of 21 autosomal STR markers in a mixed race Peruvian population applied to forensic practice, Rev. Espanola de Medicina Leg. 45,3 (2018), 92-97. https://doi.org/10.1016/j.reml.2018.09.002

[3] C.D. Neyra, E. Delgado, C. Robles, et al., Population genetic study of a Peruvian population using human identification STRs, Int J Legal Med. 134,6 (2020), 2071-2073. https://doi.org/10.1007/s00414-020-02418-6

[4] C.D. Neyra, M. Suarez, E. Cueva, et al., Identificación genética de recién nacidos en Perú: un estudio piloto. Rev Chil Pediatr 90,1 (2019), 26-35. https://dx.doi.org/10.32641/rchped.v90i1.730

[5] C.D. Neyra-Rivera, B. Budowle, Genetic variation of different Peruvian populations using 23 autosomal STR markers. Forensic Sci Int Genet Suppl Ser 8. (2022), 259–262. https://doi.org/10.1016/j.fsigss.2022.10.055

[6] C.D. Neyra-Rivera, E. Delgado, F. Diaz, et al., Genetic study with autosomal STR markers in people of the Peruvian jungle for human identification purposes, Can Soc Forensic Sci. 54, 3 (2021), 117-138. https://doi.org/10.1080/00085030.2021.1933811

[7] C.D. Neyra-Rivera, C.S. Robles, E. Delgado, et al., Analysis of 27 STR markers of the chromosome Y in the population of the jungle from Peru, Rev. Espanola de Medicina Leg. 0377-4732 (2023). https://doi.org/10.1016/j.reml.2022.10.001

[8] C.D. Neyra-Rivera, A. Ticona, E. Delgado, et al., Allelic frequencies with 23 autosomic STRS in the Aymara population of Peru, Int J Legal Med. 135, 3 (2020):779-781. https://doi.org/10.1007/s00414-020-02448-0

[9] C.D. Neyra-Rivera, A. Ticona, E Delgado, et al., Population data of 27 Y-chromosome STRS in Aymara population from Peru, Aust J Forensic Sci. (2021)1-15. https://doi.org/10.1080/00450618.2021.1882571

[10] C.D. Neyra-Rivera, E.D. Ramos, E.G. Ingunza, et al., Analysis of 27 Y-chromosomal STR loci of the Mestizo Peruvian population, Egypt J Forensic Sci 13, 47 (2023). https://doi.org/10.1186/s41935-023-00366-x

[11] M.E. Bermejo, J. Ge, B. Budowle, C. Infante, C.D. Neyra-Rivera, Genetic study with 21 autosomal STRs in five Peruvian macro regions for human identification purposes, Legal Med 57 (2022), 1344-6223. https://doi.org/10.1016/j.legalmed.2022.102073

[12] R. Pereira, C. Phillips, N. Pinto, et al., Straightforward Inference of Ancestry and Admixture Proportions through Ancestry-Informative Insertion Deletion Multiplexing, PLoS ONE 7. (2012) e29684. doi.org/10.1371/journal.pone.0029684

[13] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics. 155 (2000) 945–959. https://doi.org/10.1093/genetics/155.2.945

[14] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics, 2005. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2658868/

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 541-548
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Tracing the African maternal origins in Brazilian quilombos

Masinda Nguidi[1,*], Filipa Simão[2], Allan Scharf[3], Dalmo Azevedo[3], Leonor Gusmão[1,*]

[1]*DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro, Rio de Janeiro, Brazil.*
[2]*Department of Forensic Science, Virginia Commonwealth University, Richmond, USA.*
[3]*Forensic DNA Laboratory (LDF), Federal University of Alagoas, Maceió, Brazil.*

** Corresponding author at: DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro - Rua São Francisco Xavier, 524. Maracanã, Rio de Janeiro (RJ) Brazil. ZIP code: 20550-900. Tel.: +55(21)2334-2183; emails to: masindanguidi@gmail.com, leonorbgusmao@gmail.com*

Abstract
Human trafficking across the Atlantic Ocean left an immeasurable mark in South America's history. In Brazil, similar to what happened in other countries, enslaved Afro-descendants began rebellion movements to escape from exploitation by creating and fleeing to isolated communities, the quilombos. Because people from quilombos aimed to remain isolated, their populations tend to preserve higher African genetic inheritance. The African ethnolinguistic and geographic origin of the maternal lineages of five quilombos from the State of Alagoas, Brazil, was investigated in the present study. The genetic composition of these communities was analyzed through mitochondrial DNA control region sequencing. From a total of 227 lineages, 141 were from the African haplogroup L. Statistically significant differences were found among all quilombos, although located in the same Brazilian State. However, shared haplotypes among the communities indicate either a common background or inter-community flux of people over time. The detection of lineages with different geographic incidences within Africa suggests a diverse origin of quilombos' ancestors.

Keywords
Afro-descendant communities, quilombo, Brazil, mitochondrial DNA.

Introduction

Like many countries in America, Brazil carries a complex history of admixture among the original native American people, European colonizers and enslaved

Africans. Currently, most urban populations in Brazil present a high European genetic input, with a variable proportion of native ancestry in different regions throughout the country, due to the massive extermination of the original people during colonial times. Brazilian population also has a significant African background, and 56.1% of the Brazilians self-identified as Black or of African descent [1]. Part of this population resides in communities (quilombos) formed by groups of Afro-descendants that escaped from slavery during colonial times, or who settled in these places after abolition. It is estimated that there are around 5.972 quilombos scattered through Brazilian territory, with 63,5% being in the Northeast region, where Alagoas State is located [1].

Since their establishment, the quilombos have remained isolated, so it is expected that African genetic and cultural traits have been preserved in these communities. However, it is known that, apart from Africans and their descendants, quilombos also harbored escapees, fugitive soldiers, outlaws, and white and indigenous minorities, all fleeing European exploitation [3, 4]. In that manner, a non-African genetic input can also be expected.

To infer to what extent African maternal ancestry was preserved in these groups, the mitochondrial DNA (mtDNA) of five quilombos from Alagoas was analyzed. The ethnolinguistic or geographic origins of the African maternal lineages was inferred through phylogenetic analyses, including previously published data from African populations.

Material and methods

This study was performed under informed consent and approved by the CEP 1.753.418 (Ethical Committee from the Federal University of Alagoas). A total of 227 samples were included, from the following communities: Pau D'Arco (n = 59), Bom Despacho (n = 59), Quilombo (n = 42), Tabuleiro dos Negros (n = 42) and Vila Santo Antonio (n = 25) (Fig. 1A). Saliva samples were collected with buccal swab and DNA was extracted by Chelex method (modified from Walsh et al. [5]). The control region sequences were obtained after an initial amplification using the Qiagen Multiplex Kit (Qiagen), followed by a purification with ExoSAP (Applied Biosystems), and the purified PCR products were sequenced with BigDye v3.1 Cycle Sequencing kit (Applied Biosystems). Separation and detection of the sequencing products were performed on an ABI 3500 Genetic analyzer (Applied Biosystems).

The mtDNA haplotypes were determined with the SeqScape v2.7 software (Thermo Fisher Scientific) in comparison to the Revised Cambridge Reference Sequence (rCRS) [6]. Haplogroups were assigned on EMPOP [7]. Genetic distances (F_{ST}) and non-differentiation probabilities (significance level of 0.005, by applying Bonferroni correction) between the quilombo communities were calculated using the Arlequin ver. 3.5.1.2 software [8]. To infer the most probable African origin of L-haplotypes, median-joining networks were constructed using the Network v10.1.0.0 software (Fluxos Technology Ltd., Colchester, UK). Available data from African populations were included in the results analyses [9-17].

Results

The haplotype diversity values found for the quilombos were variable ($0.8722 \leq HD \geq 0.9895$). All 5 communities maintain high levels of maternal African ancestry (45.8% to 78.6%), with the most frequent African haplogroups being L0a (8%), L1b (25%), L2 (11%), and L3e (10%). The main non-African contribution is of Native American origin (16.7% to 52.5%), and only eight sequences (considering the total data) were attributed to Eurasian haplogroups. Bom Despacho was the only quilombo where African maternal ancestry was not the highest but rather Native-American. Pairwise F_{ST} genetic distances between the five communities were high ($0.027 \leq F_{ST} \geq 0.187$), showing statistically significant differences ($p \leq 0.0002$), except in Pau D'Arco vs. Quilombo ($p=0.0077$) and Bom Despacho vs. Quilombo ($p=0.0072$) comparisons.

When only considering the African L-lineages, genetic distances increased in all pairwise comparisons, except between Quilombo and Bom Despacho, which showed genetic similarities ($F_{ST} = -0.013$; $p=0.779$). Shared haplotypes among L-lineages were found in comparisons between (1) Pau D'Arco and Tabuleiro dos Negros, (2) Bom Despacho, Tabuleiro dos Negros and Quilombo, and (3) Pau D'arco and Bom Despacho.

A search for shared haplotypes with populations from Africa showed diverse matches for specific lineages, in samples from regions such as: Central-West (Chad, Ghana, Ivory Coast, Niger, Nigeria and Togo), Southwest (Angola), East (Rwanda) and Southeast (Mozambique) (Fig. 1).

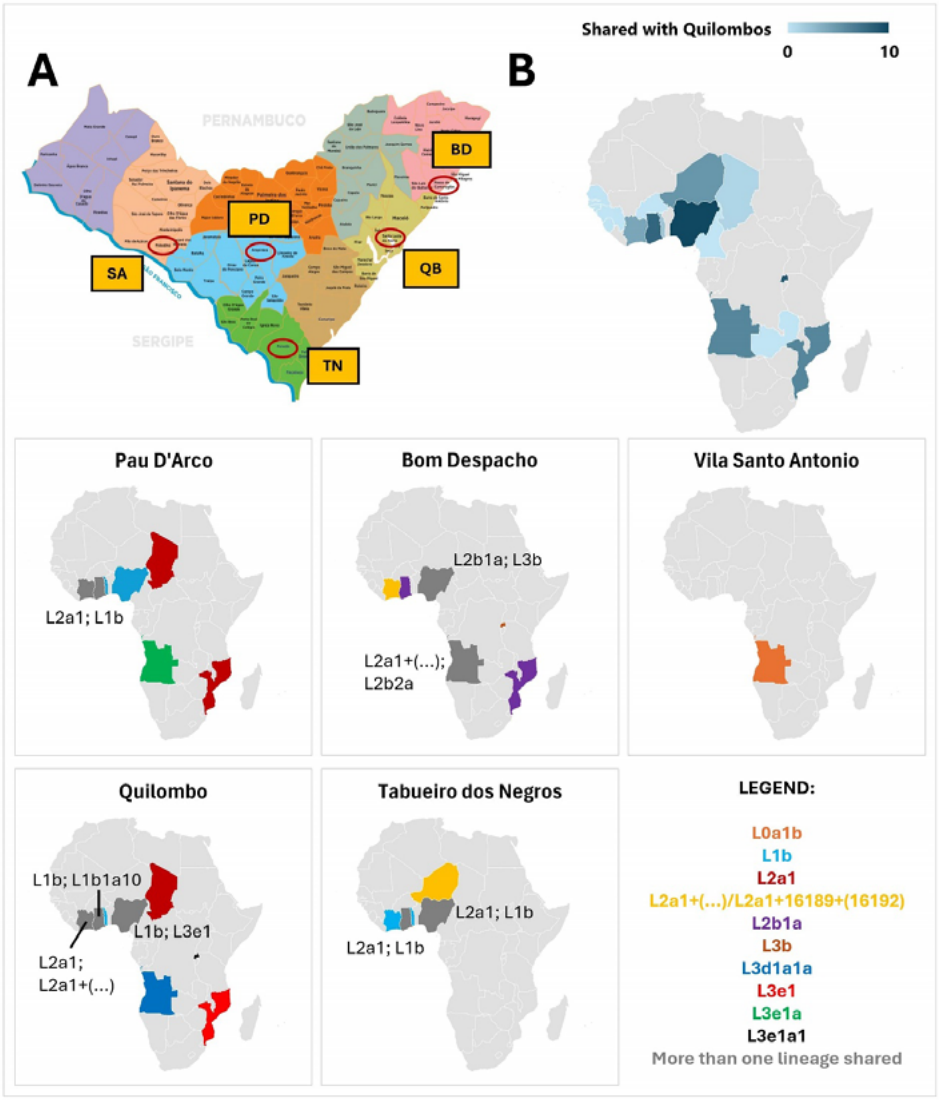


Figure 1. Maps presenting the location of the communities (A) and the number of shared haplotypes between African populations and quilombos from Alagoas (B). All the countries used in this analysis were included in Figure 1B, even the ones with no haplotype matches. The lineages present in each Quilombo that were found in Africa are shown in each box (bellow).
Notes: The abbreviations stands for: PD – Pau D'Arco; BD – Bom Despacho; SA – Vila Santo Antonio; QB – Quilombo; TN – Tabuleiro dos Negros;

Network analyses were performed (data not shown) to infer the most likely origin of the most representative African lineages found in the quilombos (Table 1).

Table 1. Results from phylogenetic analyses (Network) based on the mtDNA entire control region of the most frequent African lineages found in the quilombos and their possible origin in Africa.

	Quilombola Community				
Lineages	Pau D'Arco	Bom Despacho	Quilombo	Tabuleiro dos Negros	Vila Santo Antônio
L0a1b	-	Rwanda and Zambia	-	Mozambique	Angola, Zambia, Mozambique and Rwanda
L1b	Nigeria, Togo, Ghana and Ivory Coast	Nigeria and Zambia	Nigeria, Ghana, Togo and Zambia	Nigeria, Ghana, Togo and Ivory Coast	Probably derived from a Pau D'Arco lineage
L2a1/ L2a1+16189+(16192)	Ivory Coast, Niger and Mozambique	Angola, Zambia, Ghana, Ivory Coast and Nigeria	Ghana, Burkina Faso, Ivory Coast, Niger and Gambia	Angola, Zambia, Ghana, Ivory Coast, Chad and Nigeria	-
L3e1a	Zambia and Angola	Zambia and Angola	Zambia, Angola, Mozambique and Rwanda	-	-

Note: Analyses were performed disregarding the InDels: 16193.xC, 309.xC, 315.xC, 522-524 insertions/deletions and 573.xC.

Discussion

The geographical isolation of quilombos is reflected in their low maternal genetic diversity, when comparing to the high diversity levels commonly found in admixed populations in Brazil (ranging between 0.96-0.98). Although shared haplotypes were found among quilombos, most pairwise comparisons showed that their maternal composition is distinct. The statistically significant differences were reinforced when only the African lineages were analyzed. Such results could be indicative that quilombos with similar origins have differentiated over time, as they were successful in getting isolated in separate regions of the Alagoas state.

The presence of shared haplotypes of specific lineages between quilombos and populations from Africa (Fig. 1) showed that the maternal lineages came mostly from the Central-West and Southwest regions. However, most lineages in the quilombos had a broad distribution within Africa, preventing their tracking to a specific geographic region. Surprisingly, Rwanda was the second country with most matches with the quilombos being only behind Nigeria. This is an unexpected result since the country was not on the Transatlantic route of traffic of enslaved people to South America.

When digging deeper into the maternal origins of the most frequent African lineages in the quilombos, the networks constructed evidenced once again the high influence of countries from the Central-West and Southwest regions on Brazilian quilombos. Despite being disseminated across the continent, L2a and L3e lineages in quilombos could be traced to a specific African regional origin (Table 1), where similar haplotypes were found. A regional pattern of origin was also found in the L0a1b and L1b lineages, for Bom Despacho, Pau D'Arco and Tabuleiro dos Negros. However, it was not possible to define the ethnic origin of most lineages, due to the lack of information on ethnolinguistic affiliation of the reference samples.

Conclusion

Despite the statistically significant differences in the maternal genetic composition among quilombos, the presence of shared haplotypes suggests a common genetic background in their establishment. The haplotype sharing can also be the result of gene flow during or after the establishment of these communities. Nonetheless, such event does not appear to have been strong enough to erase genetic differentiation, most likely driven by founder events and genetic drift due to low population size.

The detection of lineages from different African regions points to a broad rather than a single geographic origin of the quilombos' genetic background. During the enslavement of Africans through the Atlantic, people from different ethnicities were deliberately brought together to difficult communication and avoid revolts during the crossings. Although the presence of lineages from across various regions of the African continent is plausible, it was not possible to attribute an ethnic affiliation to these lineages.

Acknowledgments

M.N. is supported by CAPES, Brazil. D.A. is supported by FAPEAL, Brazil. L.G. is supported by FAPERJ, Brazil (CNE-2022) and CNPq, Brazil (ref. 315007/2023-0*).*

Conflict of interest statement

None.

References

[1] IBGE/PNAD Contínua, 2021.
[2] F. Dos S. Gomes, Mocambos e Quilombos uma história do campesinato negro no Brasil, Coleção Agenda Brasileira, São Paulo, 2015.
[3] K. Munanga, Origem e histórico do quilombo na África, Revista USP, São Paulo, 28 (1996) 56–63.
[4] J. J. Reis, Quilombos e revoltas escravas no Brasil, Revista USP, São Paulo, 28 (1995) 14–39.
[5] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, Biotechniques, 10 (1991) 506-513.
[6] R. M. Andrews et al., Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nature genetic, 23 (1999) 147.
[7] EMPOP (https://empop.online).
[8] L. Excoffier, H. E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Molecular ecology resources, 3 (2010) 564-567.
[9] M. Nguidi et al., Impact of patrilocality on contrasting patterns of paternal and maternal heritage in Central-West Africa. Scientific Reports, 14 (2024).
[10] 1000 Genome Project.
[11] T.M. Göbel et al., Mitochondrial DNA variation in Sub-Saharan Africa: Forensic data from a mixed West African sample, Côte d'Ivoire (Ivory Coast), and Rwanda, Forensic Science International: Genetics, 44 (2020).
[12] L. Fendt et al., MtDNA diversity of Ghana: a forensic and phylogeographic view, Forensic Science International: Genetics, 6 (2012) 244–249.

[13] M. Čížková et al., Genetic Structure of the Western and Eastern African Sahel/Savannah Belt and the Role of Nomadic Pastoralists as Inferred from the Variation of D-Loop Mitochondrial DNA Sequences, Human biology, 89 (2017) 281-302.

[14] M. Carvalho et al., Paternal and maternal lineages in Guinea-Bissau population, Forensic Science International: Genetics, 5 (2011) 114–116.

[15] M. Cerezo et al., Comprehensive Analysis of Pan-African Mitochondrial DNA Variation Provides New Insights into Continental Variation and Demography, Journal of Genetics and Genomics, 43 (2016) 133-143.

[16] C. Barbieri et al., Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia, European Journal of Human Genetics, 21 (2013) 430–436.

[17] C. Barbieri et al., Migration and Interaction in a Contact Zone: mtDNA Variation among Bantu-Speakers in Southern Africa, PLoS ONE, 9 (2014).

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 549-552
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Allelic inconsistencies between commercial kits of autosome STR markers: technical-legal implications and recommendatios

Izquel Sánchez[1], Libardo Mendoza[1], Aníbal Gaviria[2], Ana Karina Zambrano[3], Diana Aguirre[1], Juan Builes[1,*]

[1]*Laboratorio Genes SAS, Medellín, Colombia*

[2]*Laboratorio de Genética Molecular, Centros Médicos Especializados, Cruz Roja Ecuatoriana, Quito Pichincha, Ecuador*

[3]*Centro de Investigación Genética y Genómica, Facultad de Ciencias de la Salud Eugenio Espejo, Universidad UTE, Quito Pichincha, Ecuador*

** Corresponding author at: Carrera 48 # 10-45 Cons. 611, Medellín, Colombia, jjbuilesg@laboratoriogenes.com*

Abstract

Commercially there are several A-STRs kits that share many of their markers. In this work we studied the concordance of the profiles obtained with A-STRs kits from different commercial companies in 31,309 cases of biological paternity research from Genes SAS Laboratory (Colombia) and from the Laboratory of Molecular Genetics of the Ecuadorian Red Cross (Ecuador), from July 2019 to February 2024. We found 96 discordances distributed in 13 markers among the genetic profiles obtained with the different commercial kits, D19S433 (23), D16S539 (21) and D8S1179 (20) presented more events. The other hand, 18 cases presented markers whit alleles outside the range of their allelic ladder. It is concluded that both laboratories that investigate paternity tests, complex biological relationships and forensic cases using this type of STR kits, as well as commercial companies, are obliged to know and communicate the behaviour of the genetic markers used in these kits, thus avoiding erroneous interpretations of this type of events. Likewise, commercial companies are recommended to make adjustments in the formulation of the kits in order to achieve perfect concordance of genetic profiles in individuals. Besides it is important to include the polymorphisms found in different populations to make the kits more inclusive.

Keywords

A-STRs, DNA Profiles, Allelic inconsistencies, Absent alleles, Paternity Testing, DNA typing.

Introduction

Laboratory Genes SAS and the Laboratory of Molecular Genetics of the Ecuadorian Red Cross are leaders in Colombia and Ecuador, respectively, in human identification and the establishment of Biological Relationships using DNA markers, mainly in the study of Biological Paternity. Both laboratories are ISO 9001:2015 certified and accredited by ISO/IEC 17025:2017. To investigate paternity testing both laboratories use A-STRs markers, which is the currently accepted methodology. Commercially there are several kits of A-STRs, which share many of their markers with each other. Genes and the Laboratory of Molecular Genetics of the Ecuadorian Red Cross use the VeriFiler Express kit (Thermo Fisher) to initially analyze paternities cases, and the PowerPlex Fusion kit (Promega) to perform confirmations of exclusion cases; either paternity or biological relationships and in cases with the presence of apparent mutations. In the daily activity of these verifications, inconsistencies have been detected in some markers between the VeriFlier and the PowerPlex, which led us to a more detailed analysis of the performance of these kits. The objective of this work was to evaluate the concordance of the genetic profiles of the Genes SAS laboratory and the Laboratory of Molecular Genetics of the Ecuadorian Red Cross obtained with A-STRs kits from two different commercial houses, to establish possible technical-legal implications and suggest some recommendations for the non concordant results.

Materials and Methods

In this work we studied the concordance of the profiles obtained with A-STRs kits from different commercial companies in 22.114 cases of biological paternity research from Genes SAS and 9,195 cases from the Laboratory of Molecular Genetics of the Ecuadorian Red Cross.

Genetic profiles of 11.560 cases from 31.309 paternity tests attended in both laboratories from July 2019 to February 2024, initially studied with the VeriFiler Express kit (Thermo Fisher) and subsequently verified with their counter-samples using the PowerPlex Fusion kit (Promega), were comparatively analyzed to confirm paternity exclusion or verify any new paternal and/or maternal mutation.

Results and Discussion

We found 96 discordances distributed in 13 markers among the genetic profiles obtained with the different commercial kits in the 11,560 cases verified, 94 cases presented absence of an allele in a marker for one of the kits, D19S433 (23), D8S1179 (20) and D16s539 (19) presented more events, and two markers presented different genotypes. The other hand, 18 cases presented markers whit alleles outside the range of their allelic ladder (Fig. 1-2).

According to the literature [1-6], these findings may be due to differences in primer sequences between kits for the same marker, which are complementary to the flanking regions of the DNA segment to be amplified, together with a possible mutation in these complementary sequences preventing the total or partial amplification of the sequence of interest. These events are conventionally referred to as null or silent alleles, for these particular cases we have called them only as absent, since it has been possible to establish the real alleles with an alternate case.

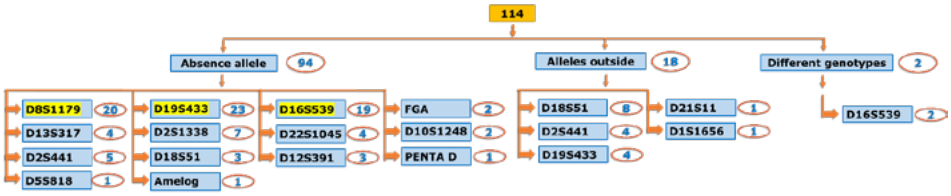


Figure 1. Classification and number of inconsistencies per marker

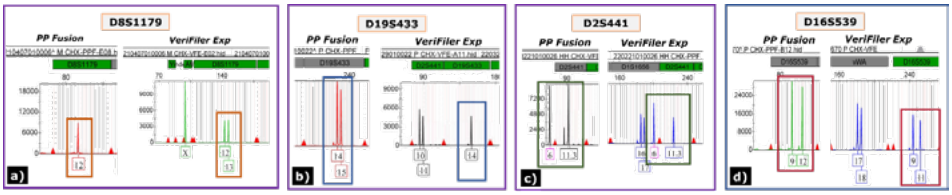


Figure 2. Examples of allelic inconsistencies found in this study: a) and b) missing alleles; c) different genotypes and d) alleles outside their allelic ladder

Conclusion

Laboratories that investigate paternity tests, complex biological relationships and forensic cases using these types of STR kits, as well as commercial companies, are

obliged to know and communicate the behaviour of the genetic markers used in these kits, thus avoiding erroneous interpretations of this type of events. Likewise, commercial companies are recommended to make adjustments in the formulation of the kits in order to achieve perfect concordance of genetic profiles in individuals. Besides it is important to include the polymorphisms found in different populations to make the kits more inclusive. This work shows the importance of establishing the frequency of silent or null alleles for each marker in all kits based on the reference population where it is being used.

Conflict of interest statement

None.

References

[1] B. Budowle, A. Masibay, S.J. Anderson et al., STR primer concordance study, Forensic Sci. Int. 124 (2001) 47–54.

[2] M. Delamoye, C. Duverneuil, K. Riva et al., False homozygosities at various loci revealed by discrepancies between commercial kits: implications for genetic databases, Forensic Sci. Int. 143 (2004) 47–52.

[3] T.M. Clayton, S.M. Hill, L.A. Denton et al., Primer binding site mutations affecting the typing of STR loci contained within the AMPFlSTR® SGM Plus™ kit, Forensic Sci. Int. 139 (2004) 255–259.

[4] N. Vanderheyden, A. Mai, A. Gilissen et al., Identification and sequence analysis of discordant phenotypes between AmpFlSTR SGM Plus™ and PowerPlex® 16, Int. J. Legal Med. 121 (2007) 297–301.

[5] C. Leibelt, B. Budowle, P. Collins et al., Identification of a D8S1179 primer binding site mutation and the validation of a primer designed to recover null alleles, Forensic Sci. Int. 133 (2003) 220–227.

[6] M.S. Nelson, E.N. Levedakou, J.R. Matthews et al., Detection of a Primer-Binding Site Polymorphism for the STR Locus D16S539 Using the Powerplex® 1.1 System and Validation of a Degenerate Primer to Correct for the Polymorphism, J. Forensic Sci.47 (2002) 345–349.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 553-559
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Sequence-based North-East Italy population data of five autosomal STRs exclusively included in the ForenSeq™ DNA Signature Prep Kit

Stefania Turrina[1*], Chiara Saccardo[1], Giulia Soldati[1], Dario Raniero[1], Riccardo Saccà[1], Rachele Turrini[1], Domenico De Leo[1], Francesco Ausania[1]

[1]*Department of Diagnostics and Public Health, Section of Forensic Medicine, Forensic Genetics Lab, University of Verona, Verona, Italy.*

**Corresponding author at: Policlinico G.B. Rossi, P.le L.A. Rossi, 37134 – Verona (Italy), stefania.turrina@univr.it, telephone number: +39 0458027622*

Abstract

The inability to share between capillary electrophoresis (CE) and massively parallel sequencing (MPS) of some short tandem repeats (STRs) included exclusively in the MPS kits, associated with the limited availability of MPS technology in forensic genetics laboratories, does not allow to have for these STRs an adequate allele frequency database, precluding the possibility of performing appropriate statistical calculations and the use of these loci as additional markers in the different forensic casework.

To contribute to overcoming this issue, in this study, DNA samples of 255 individuals native to North-East Italy were genotyped for the first time, with the five additional autosomal STRs (auSTRs) (D4S2408, D6S1043, D9S1122, D17S1301, and D20S482) included in the ForenSeq™ DNA Signature Prep Kit.

When data analysis moved from length-based to sequence-based alleles, it was observed for the five auSTRs an overall increment in the number of alleles from 42 to 55. In particular, sequence variations affected forensic statistical parameters for D4S2408, D9S1122, and D20S482.

Keywords

Massively parallel sequencing (MPS), ForenSeq™ DNA Signature Prep Kit, Short Tandem Repeats (STRs), Allele frequency.

Introduction

For over a decade, the MPS has been introduced in forensic genetics due to its informative potential on genetic markers.

One of the forensic kits used for MPS is the ForenSeq™ DNA Signature Prep Kit (FSSP), for which various validation studies have been conducted that evidence its potential advantages in forensic applications arising from the detection of sequence variations in the target STRs and SNPs as well as in their flanking regions that affect positively the forensic parameters of these markers [1-5].

Out of the 27 autosomal STRs (auSTRs) comprised in DNA primer Mix A (DPMA) of the FSSP kit, five of them (D4S2408, D6S1043, D9S1122, D17S1301, and D20S482) are additional autosomal STRs that are not commonly included in the STR kits analyzed by capillary electrophoresis (CE). To enable their use as additional markers and include them in biostatistical calculations, it is necessary to acquire extensive data on allele frequencies and forensic parameters in different populations in order to provide valid support in solving various forensic contexts.

Here are reported sequence variations, allele frequencies, and forensic parameters for these five auSTRs evaluated in North-Eastern Italian subjects.

Materials and methods

Anonymized buccal swab samples were collected from 255 unrelated autochthonous individuals from North-East Italy after obtaining informed consent and ethical approval from the University of Verona's Research Ethics Committee (protocol code: CARU-12/2020).

DNA was extracted using the QIAamp® DNA Mini Kit (Qiagen) and quantified with the Qubit® dsDNA HS Assay Kit on Qubit® 2.0 Fluorometer (Thermo Fisher Scientific) and normalized to 1 ng/µl.

The DNA Primer Mix A (DPMA) of the ForenSeq™ DNA Signature Prep Kit (FSSP) (Verogen) [6] was used for MPS library preparation. Sequencing was performed on the MiSeq FGx™ Forensic Genomics System [7], and data analysis was carried out by Universal Analysis Software (UAS) v1.2 (Verogen), applying default parameters. Each DNA sample was sequenced twice, and only the genetic profiles with a total read count equal to or greater than 85,000 were considered to ensure the reproducibility of the data.

The FASTQ files automatically generated by UAS were reanalyzed using the STRait Razor v3.0 (SRv3) pipeline [8] in its default configuration (ForenSeqv1.27.config) to compare STR genotypes and isoalleles calls with UAS's ones. Furthermore, SRv3 has been used to reveal sequence variations at the single allele in the heterozygous genotype and in the flanking regions of target STRs, which were not detectable by the available UAS version.

For sequence variations not identified by UAS, the comprehensive allelic nomenclature was assigned directly by SRv3. However, when also SRv3 did not recognize the sequence variation because it was not included in its string-matching database since it was rare or not yet described, the nomenclature was attributed manually after having aligned the sequence string with the human reference sequence reported by the UCSC Genome Browser (GRCh38) to identify the chromosomal location of the mutation and the involved SNP.

Regarding population genetics analysis, allele frequencies, observed heterozygosity (Ho) and expected heterozygosity (He), match probability (MP), typical paternity index, power of exclusion (PE), and power of discrimination (PD), were determined based on length and sequence of each allele detected for each of five auSTRs. Moreover, the Hardy-Weinberg equilibrium (HWE) test was performed on length- and sequence-based alleles.

Results and Discussion

Both bioinformatic pipelines identified the presence of isoalleles in homozygous genotypes (alleles with same length but different sequences) at loci D4S2408 (three isoalleles 9), D9S1122 (three isoalleles 11, nineteen isoalleles 12, and nineteen isoalleles 13), and D20S482 (one isoallele 13, nine isoalleles 14 and one isoallele 16) (Table 1).

Furthermore, SRv3 revealed the presence of a sequence mutated compared to the reference one shared by the two alleles of some homozygous genotypes at the D6S1043 (6 genotypes), D9S112 (21 genotypes) and D20S482 (1 genotype) loci. SRv3 also detected sequence variations compared to the reference sequence for a single allele in several heterozygous genotypes at D4S2408 (5 sequence variations), D6S1043 (131 sequence variations), D9S112 (184 sequence variations), D17S1301 (1 sequence variation), and D20S482 (24 sequence variations). A total of 456 sequence variations were detected, of which 401 were only revealed by SRv3, which assigned a correct nomenclature to 395 of these already reported in the literature

[1-5]. For the remaining six sequence variations, SRv3 was unable to provide the allelic nomenclature tagging them with NA, probably because they were not present in its sequence string-matching database due to the rarity (<0.01).

In particular, three of the six sequence variations, never described before this study was performed, have been concerned D6S1043 where the SNP rs1251963869 was identified in the reverse flanking region of an allele 12, and D9S1122 where the SNPs rs4281164 and rs1020194359, were identified in the repeat motif of an allele 8 and an allele 12, respectively (Table 1). Overall, for the five auSTRs, moving from the length-based to sequence-based allele calls, it was observed a rise in the number of typologies of detected alleles from 42 to 55, where the higher increases were found for D9S1122 (from 9 to 14), D20S482 (from 8 to 12), and D6S1043 (from 12 to 15).

None of the five auSTRs deviated from the expected value ($p>0.05$) for the Hardy-Weinberg equilibrium when the tests were performed based on alleles length and sequence.

Concerning the forensic parameters determined for the five auSTRs, it was possible to observe how parameters' values are positively influenced by the detection of isoalleles compared to the estimation based only on amplicons' length, which becomes even more significant due to the detection of a mutated sequence shared by the two alleles in homozygous genotypes and of a sequence variation involving one of the two alleles in heterozygous genotypes (Table2).

In particular, regarding the PD values, although it was observed that for the population studied, these are already high for all five auSTRs when determined based on the length of the alleles, more significant increases were found when the values were calculated based on allelic sequence variations. For example, for the loci D20S482 and D9S1122, a greater number of sequence variations were found that led to an increase in the number of alleles of the two loci and, as a consequence, of their PD values that became 0.903 and 0.946, respectively.

Table1. Detected allele sequences variations

auSTR	Allele	Comprehensive Nomenclature			SNP Chromosomal location	rsID number
		Sequence variations	Rare sequence variations (<0.01)	Never described sequence variations		
D4S2408	9	D4S2408 [CE 9]-GRCh38-chr4:31302792-31302841 ATCT GTCT [ATCT]7			Chr4:31302802 A>G	rs12501500
D6S1043	8		[NA] D6S1043 [CE 8]-GRCh38-chr6:91740160-91740292 [ATCT]8			
	12			**[NA] D6S1043 [CE 12]-GRCh38-chr6:91740160-91740292 [ATCT]12**	Chr6:91740278 A>C	rs1251963869
	15	D6S1043 [CE 15]-GRCh38-chr6:91740160-91740292 ATCT ATGT [ATCT]13			Chr6:91740231 C>G	rs546148333
	17	D6S1043 [CE 17]-GRCh38-chr6:91740160-91740292 [ATCT]5 ATGT [ATCT]11			Chr6:91740247 C>G	rs11965429
	18	D6S1043 [CE 18]-GRCh38-chr6:91740160-91740292 [ATCT]5 ATGT [ATCT]12			Chr6:91740247 C>G	rs11965429
	19	D6S1043 [CE 19]-GRCh38-chr6:91740160-91740292 [ATCT]5 ATGT [ATCT]13			Chr6:91740247 C>G	rs11965429
	20	D6S1043 [CE 20]-GRCh38-chr6:91740160-91740292 [ATCT]5 ATGT [ATCT]14			Chr6:91740247 C>G	rs11965429
	21	D6S1043 [CE 21]-GRCh38-chr6:91740160-91740292 [ATCT]5 ATGT [ATCT]15			Chr6:91740247 C>G	rs11965429
	21	D6S1043 [CE 21]-GRCh38-chr6:91740160-91740292 [ATCT]6 ATGT [ATCT]14			Chr6:91740251 C>G	rs191127136
D9S1122	7		[NA] D9S1122 [CE 7]-GRCh38-chr9:77073809-77073880 [TAGA]7			
	8			**[NA] D9S1122 [CE 8]-GRCh38-chr9:77073809-77073880 TAGA TCGA [TAGA]6**	Chr9:77073831 A>C	rs4281164
	9	D9S1122 [CE 9]-GRCh38-chr9:77073809-77073880 TAGA TCGA [TAGA]7			Chr9:77073831 A>C	rs4281164
	11	D9S1122 [CE 11]-GRCh38-chr9:77073809-77073880 TAGA TCGA [TAGA]9			Chr9:77073831 A>C	rs4281164
	12	D9S1122 [CE 12]-GRCh38-chr9:77073809-77073880 TAGA TCGA [TAGA]10			Chr9:77073831 A>C	rs4281164
	12			[NA] D9S1122 [CE 12]-GRCh38-chr9:77073809-77073880 TAGG [TAGA]11	Chr9:77073829 A>G	rs1020194359
	13	D9S1122 [CE 13]-GRCh38-chr9:77073809-77073880 TAGA TCGA [TAGA]11			Chr9:77073831 A>C	rs4281164
	14	D9S1122 [CE 14]-GRCh38-chr9:77073809-77073880 TAGA TCGA [TAGA]12			Chr9:77073831 A>C	rs4281164
D17S1301	8		**[NA] D17S1301 [CE 8]-GRCh38-chr17:74684846-74684918 [AGAT]8**			
D20S482	13	D20S482 [CE 13]-GRCh38-chr20:4525674-4525771 [AGAT]13 4525680-T			Chr20:4525680 C>T	rs77560248
	14	D20S482 [CE 14]-GRCh38-chr20:4525674-4525771 [AGAT]14 4525680-T			Chr20:4525680 C>T	rs77560248
	15	D20S482 [CE 15]-GRCh38-chr20:4525674-4525771 [AGAT]15 4525680-T			Chr20:4525680 C>T	rs77560248
	16	D20S482 [CE 16]-GRCh38-chr20:4525674-4525771 [AGAT]16 4525680-T			Chr20:4525680 C>T	rs77560248

Table 2. Population forensic parameters for auSTRs included in ForenSeq™ DNA Signature Prep Kit

auSTR	Typing		Number of Alleles	Hobs	Typical Paternity Index	Match Probability	Power of Exclusion	Power of Discrimination
D4S2408	length-based		6	0,758	2,361	0,111	0,577	0,889
	sequence-based	UAS	7	0,761	2,500	0,109	0,599	0,891
		SRv3	7	0,766	2,500	0,104	0,599	0,896
D6S1043	length-based		12	0,804	2,452	0,065	0,592	0,935
	sequence-based	UAS	12	0,804	2,452	0,065	0,592	0,935
		SRv3	15	0,805	2,452	0,065	0,592	0,935
D9S1122	length-based		9	0,693	1,656	0,150	0,425	0,850
	sequence-based	UAS	12	0,744	3,542	0,130	0,712	0,870
		SRv3	14	0,835	3,542	0,054	0,712	0,946
D17S1301	length-based		7	0,643	1,262	0,187	0,296	0,813
	sequence-based	UAS	7	0,643	1,262	0,187	0,296	0,813
		SRv3	7	0,643	1,262	0,187	0,296	0,813
D20S482	length-based		8	0,700	1,433	0,131	0,357	0,869
	sequence-based	UAS	11	0,718	1,635	0,116	0,419	0,884
		SRv3	12	0,742	1,656	0,097	0,425	0,903

Conclusion

Although several validation studies have been carried out to promote its use, population data on some markers included exclusively in MPS kits is still insufficient to allow their application as additional markers in some forensic circumstances.

In this study, were estimated, for the first time in a population from North East Italy, allelic frequencies and forensic parameters based on the allele length and sequence of five auSTRs (D4S2408, D6S1043, D9S1122, D17S1301 e D20S482) included in ForenSeq™ DNA Signature Prep Kit.

Based on the findings provided by alleles' length analysis, it can be inferred that the five auSTRs are characterized by high PD values that further increase when the estimation is made on sequence-based allelic variations.

However, it should be noted that UAS tends to underestimate the presence of allele sequence variations to the detriment of forensic parameters, failing to highlight them when manifested as a mutated sequence shared by both alleles of homozygous genotypes and when affecting alleles of heterozygous genotypes, making it necessary to process the raw data with another bioinformatic system to compensate for this lack.

Conflict of interest statement

None.

References

[1] L. Devesse, D. Ballard, L. Davenport, et al., Concordance of the ForenSeq™ system and characterisation of sequence-specific autosomal STR alleles across two major population groups. Forensic Sci Int Genet. 34 (2018) 57-61. doi.org/10.1016/j.fsigen.2017.10.012.

[2] K.B. Gettings, L.A. Borsuk, C.R. Steffen, et al., Sequence-based U.S. population data for 27 autosomal STR loci. Forensic Sci Int Genet. 37 (2018) 106-115. doi.org/10.1016/j.fsigen.2018.07.013.

[3] L. Devesse, L. Davenport, L. Borsuk, et al., Classification of STR allelic variation using massively parallel sequencing and assessment of flanking region power. Forensic Sci Int Genet. 48 (2020) 102356. doi.org/10.1016/j.fsigen.2020.102356.

[4] J. Hoogenboom, N. Weiler, L. Busscher, et al., Advancing FDSTools by integrating STRNaming 1.1. Forensic Sci Int Genet. 61 (2022) 102768. doi.org/10.1016/j.fsigen.2022.102768.

[5] M.M. Agudo, H. Aanes, M. Albert, et al., An overview of autosomal STRs and identity SNPs in a Norwegian population using massively parallel sequencing. Forensic Sci Int Genet. 71 (2024) 103057. doi.org/10.1016/j.fsigen.2024.103057.

[6] Verogen (2022) ForenSeq™ DNA Signature Prep Reference Guide. Document # VD2018005 Rev. D.

[7] Verogen (2021) MiSeq FGx™ Sequencing System Reference Guide. Document # VD2018006 Rev. F

[8] A.E. Woerner, J.L. King, B. Budowle, Fast STR allele identification with STRait Razor 3.0, Forensic Sci Int Genet. 30 (2017) 18-23.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 561-569
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Analysis of 23 autosomal STRs in coastal, sierra and jungle populations from Peru for human identification purposes

Sonia Guillen Ramirez[1,2], Carlos Neyra-Rivera[3*], Ruth García-de-la-Guarda[1], Bruce Budowle[4]

[1]*Laboratorio de Microbiología Molecular y Biotecnología, Facultad de Ciencias Biológicas, Universidad Nacional Mayor de San Marcos, Lima, Peru.*

[2]*Dirección de Criminalística, Policía Nacional del Perú, Lima, Peru.*

[3]*Facultad de Ciencias de la Salud, Universidad Tecnológica del Perú, Lima, Peru.*

[4]*Department of Forensic Medicine, University of Helsinki, Helsinki, Finlandia.*

** Corresponding author at: Av. Arequipa 265, Lima 15046, Lima-Peru, carlosdavidmp@outlook.es*

Abstract

Population samples from three regions in Peru were analyzed for 23 autosomal STR markers to develop allele frequency data, to assess the significance of forensic genetic results, and to support forensic practices. The STR profiles of 524 unrelated Peruvian mestizo individuals (coastal (n=309), sierra (n=181) and jungle (n=34)) were generated and analyzed. The Arlequin v3.5.2.2 software was used to determine the genetic distances (Fst) of the studied populations. No detectable population genetic subdivision was observed among the population sample of Peruvian mestizo compared with populations from Ecuador, Bolivia, the Aymara people and the macroregions of Peru. In contrast, when comparing the populations of the coast, sierra and jungle of Peru with other Peruvian populations, there were differences with the sample populations of the departments of Amazonas, Madre de Dios, and the Ashaninkas and the populations of the coast and sierra of Peru with the population of Loreto.

Keywords

Autosomal STR, mestizo, Peru.

Introduction

The Andes Mountain range divides Peru into three regions - coastal, sierra and jungle. Climate changes, high altitude and interaction with other ethnic groups

have sculpted genetic variation and genetic expression in Peruvians. The history and migration of the Peruvian population show a predominantly Native American ancestry, followed by admixture with European and African ancestry. In contrast, there are some isolated Native American populations that experienced minimal external gene flow. Therefore, it is important to study the genetic profiles of these modern mestizo populations to better understand their underlying genetic variation in order to increase current studies [1-12]. The objective of this study was to analyze 23 autosomal STR markers and obtain allele frequency data on populations residing in three regions of Peruvian, assess the significance of forensic genetic results, and to support forensic practices.

Material studied, methods, techniques

This research has been approved by the Ethics Committee N°042-2023-CBE-FCB-UNMSM of the Universidad Nacional Mayor de San Marcos (Peru), the samples of 524 unrelated Peruvian mestizo individuals (coastal (n=309), sierra (n=181) and jungle (n=34) were collected. The samples ranged from blood, saliva, semen, sweat, feces, or urine, body parts like hair, bone, or teeth, and substrates such as sterile swabs or FTA cards containing blood or saliva samples. For DNA extraction from blood samples on FTA cards and filter paper, the FTA method was used. For samples collected on saliva, semen, sweat, feces, or urine, body parts like hair, bone, or teeth, and substrates such as sterile swabs the DNA extraction was performed using the PrepFiler Express™ Kit on the automated AutoMate Express Forensic DNA Extraction System following the manufacturer's instructions. Quantification of DNA was carried out using the Quantifiler™ Trio DNA kit according to the manufacturer's instructions. The genetic markers were amplified by PCR multiplex using the Verifiler Kit (Thermo Fisher Scientific), with 2 ng input DNA per sample in a reaction volume of 25 µL. The amplifiied products were separated and detected by capillary electrophoresis on the Applied Biosystems™ 3500 Genetic Analyzer. Subsequently, the data were imported into the GeneMapper™ ID-X v1.6 genetic typing software (Thermo Fisher Scientific). Discrimination power (PD), polymorphic information content (PIC), exclusion probability (PE), and observed heterozygosity (Ho) were calculated with the Power Stats V1.2 software [13]. The Arlequin v3.5.2.2 software was used to determine the genetic distances (Fst) of the studied populations [14]. The genetic distances were used to generate a UPGMA tree using the MEGA 6.06 program [15].

Results and Discussion

In the combined Peruvian, as well as the individual coastal, sierra and jungle Peruvian populations, the highest PIC values were for PENTA E (0.8875, 0.8881, 0.8907, and 0.8339, respectively) and FGA (0.8498, 0.8592, 0.8513 and 0.8390, respectively). In general, all markers for the combined Peruvian population and the individual populations from the coastal, sierra and jungle regions of Peru had PIC values >0.5, with the exception of the D2S441 marker for the combined Peruvian, sierra and jungle populations. The highest PD values for the combined Peruvian, coastal and sierra populations were observed in the markers PENTA E (0.9680, 0.9756 and 0.9787, respectively), FGA (0.9576, 0.9673 and 0.9608 respectively) and D6S1043 (0.9494, 0.9642 and 0.9618 respectively), while the jungle population had the highest PD in the markers D19S433 (0.9325), FGA (0.9446) and PENTA E (0.9498). The STRs that did not meet Hardy Weinberg expectations were D1S1656, PENTA D and in the combined Peruvian population, D1S1656, D2S441 and TPOX in the coastal population, D7S820 in the sierra population and D3S1358, D2S1338, D21S11, D13S317 and D1S1656 in the jungle population. After applying the Bonferroni correction (significance value p=0.0022) only the D1S1656 marker still significantly departed from expectations in the combined Peruvian population and coastal population (Table 1).

Analyzing the Reynolds distance values (Fst), the combined population sample of Peru yielded Fst values <0.01 when compared with the populations of Ecuador [16], Bolivia [17], departments located in the macroregions of northern, central, southern, eastern Peru, Lima – Callao [11] and the Aymara population of Peru [8]. On the other hand, the combined population sample of Peru yielded Fst values >0.01 when compared with the populations of Mexico [18], Hispanic American population [19], Loreto, Amazonas [6] and Ashaninka from Peru [12], demonstrating that there is detectable population substructure between the population sample of Peru and the indicated populations (Figure 1 and Table 2).

Table 1. Genetic parameters of the mestizo peruvian sample population, coastal, sierra and jungle population

MESTIZO PERUVIAN POPULATION																							
Genetic parameters	D3S1358	vWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	D10S1248	D1S1656	D12S391	D2S1338	PENTA E	D6S1043	PENTA D
PD	0,8376	0,8488	0,8772	0,8643	0,7979	0,9091	0,9373	0,9034	0,7418	0,9325	0,8267	0,9576	0,7665	0,8267	0,9313	0,8695	0,8421	0,9448	0,9217	0,8483	0,9680	0,9494	0,9308
PIC	0,6379	0,6376	0,7043	0,6593	0,5701	0,7464	0,8117	0,7645	0,4955	0,7860	0,6069	0,8498	0,5279	0,6169	0,7832	0,6745	0,6239	0,8299	0,7638	0,6579	0,8875	0,8273	0,7805
PE	0,3756	0,4504	0,5317	0,4095	0,2965	0,5497	0,6328	0,5813	0,2358	0,6504	0,3024	0,6722	0,2792	0,3503	0,6102	0,4374	0,3535	0,6658	0,5678	0,3695	0,7876	0,5713	0,5097
Het	0,6405	0,7104	0,7556	0,6871	0,6038	0,7713	0,8167	0,7664	0,5484	0,8248	0,6050	0,8379	0,5889	0,6459	0,8056	0,7054	0,6470	0,8350	0,7818	0,6373	0,8960	0,7804	0,7449
P-Value	0,0123	0,6620	0,2686	0,5665	0,0085	0,2582	0,1803	0,8277	0,0939	0,5116	0,0179	0,1192	0,0471	0,0676	0,6253	0,5876	0,3891	0,0000	0,9415	0,1376	0,1188	0,1145	0,0049
PM	0,1624	0,1512	0,1228	0,1357	0,2021	0,0909	0,0627	0,0966	0,2582	0,0675	0,1733	0,0424	0,2335	0,1733	0,0687	0,1305	0,1579	0,0552	0,0783	0,1517	0,0320	0,0506	0,0692
Het(exp)	0,6460	0,6971	0,7876	0,7003	0,6121	0,7874	0,8377	0,8613	0,9000	0,5647	0,8139	0,6562	0,8713	0,6143	0,6480	0,8130	0,7194	0,8598	0,6811	0,8530	0,7978	0,8319	0,8121
Het(obs)	0,6221	0,6794	0,8149	0,6851	0,6050	0,7748	0,7901	0,8760	0,8779	0,5553	0,7958	0,6412	0,8550	0,6069	0,6527	0,8168	0,6947	0,8263	0,6679	0,8359	0,7996	0,8168	0,7920
COASTAL POPULATION																							
Genetic parameters	D3S1358	vWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	D10S1248	D1S1656	D12S391	D2S1338	PENTA E	D6S1043	PENTA D
PD	0,8322	0,8664	0,9170	0,8670	0,7839	0,9208	0,9545	0,9622	0,8010	0,9450	0,8449	0,9673	0,7594	0,8510	0,9366	0,8841	0,8578	0,9616	0,9284	0,9491	0,9756	0,9642	0,9391
PIC	0,6068	0,6617	0,7525	0,6568	0,5595	0,7493	0,8162	0,8459	0,5565	0,8072	0,6237	0,8592	0,5397	0,6255	0,7894	0,6800	0,6490	0,8394	0,7845	0,8125	0,8881	0,8489	0,7948
PE	0,3298	0,3974	0,6282	0,3786	0,3214	0,5001	0,5682	0,7553	0,2779	0,6099	0,3786	0,7099	0,3256	0,3833	0,6717	0,4218	0,4268	0,6467	0,6654	0,6099	0,7422	0,6888	0,6467
Het	0,6311	0,6796	0,8155	0,6667	0,6246	0,7443	0,7832	0,8803	0,5890	0,8058	0,6667	0,8576	0,6278	0,6699	0,8382	0,6958	0,6990	0,8252	0,8350	0,8058	0,8738	0,8469	0,8252
P-Value	0,2807	0,3443	0,4670	0,3212	0,0056	0,2159	0,5117	0,6294	0,0044	0,4299	0,0113	0,3774	0,0547	0,0210	0,7632	0,4713	0,9039	0,0003	0,8001	0,3941	0,0105	0,0332	0,1622
PM	0,1678	0,1336	0,0830	0,1330	0,2161	0,0792	0,0455	0,0378	0,1990	0,0550	0,1551	0,0327	0,2406	0,1490	0,0634	0,1159	0,1422	0,0384	0,0716	0,0509	0,0244	0,0358	0,0609
Het(exp)	0,6623	0,7099	0,7880	0,7070	0,6103	0,7807	0,8380	0,8623	0,8977	0,6051	0,8289	0,6739	0,8739	0,6163	0,6606	0,8155	0,7264	0,8630	0,7005	0,8573	0,8108	0,8339	0,8206
Het(obs)	0,6311	0,6796	0,8155	0,6667	0,6246	0,7443	0,7832	0,8803	0,8738	0,5890	0,8058	0,6667	0,8576	0,6278	0,6699	0,8382	0,6958	0,8479	0,6990	0,8252	0,8350	0,8058	0,8252
SIERRA POPULATION																							
Genetic parameters	D3S1358	vWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	D10S1248	D1S1656	D12S391	D2S1338	PENTA E	D6S1043	PENTA D
PD	0,7879	0,8375	0,9082	0,8384	0,7811	0,9241	0,9439	0,9592	0,7031	0,9200	0,7823	0,9608	0,7790	0,8210	0,9298	0,8629	0,8052	0,9471	0,9111	0,7879	0,9787	0,9618	0,9225
PIC	0,5512	0,6113	0,7453	0,6254	0,5439	0,7671	0,8007	0,8452	0,4554	0,7515	0,5571	0,8513	0,5343	0,5879	0,7786	0,6530	0,5736	0,8199	0,7356	0,5512	0,8907	0,8436	0,7605
PE	0,3355	0,3656	0,6218	0,4570	0,2554	0,6641	0,5707	0,7740	0,1851	0,5217	0,3140	0,7184	0,2677	0,3503	0,5707	0,4056	0,3211	0,7074	0,5027	0,3355	0,7405	0,6323	0,5313
Het	0,6354	0,6575	0,8122	0,7182	0,5691	0,8343	0,7845	0,8895	0,4972	0,7569	0,6188	0,8619	0,5801	0,6464	0,7845	0,6851	0,6243	0,8564	0,7459	0,6354	0,8729	0,8177	0,7624
P-Value	0,5101	0,4291	0,0626	0,2708	0,2022	0,5555	0,1968	0,2641	0,5248	0,7801	0,4890	0,1471	0,2561	0,6059	0,2326	0,0242	0,1855	0,0777	0,7127	0,1495	0,5109	0,4161	0,0594
PM	0,2121	0,1625	0,0918	0,1616	0,2189	0,0759	0,0561	0,0408	0,2969	0,0800	0,2177	0,0392	0,2210	0,1790	0,0702	0,1371	0,1948	0,0529	0,0889	0,2121	0,0213	0,0382	0,0775
Het(exp)	0,6107	0,6719	0,7834	0,6808	0,6000	0,7986	0,8250	0,8632	0,9013	0,4952	0,7810	0,6173	0,8681	0,6130	0,6242	0,8053	0,7039	0,8591	0,6315	0,8417	0,7714	0,8288	0,7926
Het(obs)	0,6354	0,6575	0,8122	0,7182	0,5691	0,8343	0,7845	0,8895	0,8729	0,4972	0,7569	0,6188	0,8619	0,5801	0,6464	0,7845	0,6851	0,8177	0,6243	0,8564	0,7459	0,8508	0,7624
JUNGLE POPULATION																							
Genetic parameters	D3S1358	vWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	D10S1248	D1S1656	D12S391	D2S1338	PENTA E	D6S1043	PENTA D
PD	0,8080	0,8426	0,8910	0,8875	0,8287	0,8824	0,9135	0,7889	0,7215	0,9325	0,8529	0,9446	0,7612	0,8082	0,9273	0,8616	0,8633	0,9256	0,9256	0,8080	0,9498	0,9221	0,9308
PIC	0,6100	0,6399	0,7609	0,6956	0,6069	0,7229	0,8181	0,6023	0,4747	0,7991	0,6400	0,8390	0,5097	0,6373	0,7817	0,6906	0,6491	0,8303	0,7714	0,6100	0,8839	0,7894	0,7862
PE	0,1630	0,5882	0,6434	0,3928	0,3126	0,4849	0,7595	0,2146	0,2444	0,8195	0,2146	0,5882	0,2444	0,3173	0,5882	0,4849	0,3126	0,6434	0,5353	0,1630	0,8800	0,3928	0,3512
Het	0,4706	0,7941	0,8235	0,6765	0,6176	0,7353	0,8824	0,5294	0,5588	0,9118	0,5294	0,7941	0,5588	0,6214	0,7941	0,7353	0,6176	0,8235	0,7647	0,4706	0,9412	0,6765	0,6471
P-Value	0,0031	0,9890	0,1811	0,8351	0,4659	0,0515	0,0340	0,2180	0,2999	0,7172	0,9327	0,0591	0,4946	0,2486	0,0392	0,7312	0,0857	0,0395	0,4694	0,0177	0,8216	0,0687	0,1923
PM	0,1920	0,1574	0,1090	0,1125	0,1713	0,1176	0,0865	0,2111	0,2785	0,0675	0,1471	0,0554	0,2388	0,1918	0,0727	0,1384	0,1367	0,0744	0,0744	0,1920	0,0502	0,0779	0,0692
Het(exp)	0,6686	0,6980	0,8042	0,7489	0,6800	0,7665	0,8508	0,8411	0,9065	0,5237	0,8328	0,6980	0,8670	0,5953	0,6457	0,8196	0,7463	0,8196	0,7142	0,8600	0,8112	0,8266	0,8253
Het(obs)	0,4706	0,7941	0,8235	0,6765	0,6177	0,7353	0,8824	0,7647	0,9412	0,5588	0,9118	0,5294	0,7941	0,5588	0,5294	0,7941	0,7353	0,6765	0,6177	0,8235	0,7647	0,7353	0,6471

Table 2. *Average Fst values obtained between pairs of populations for 15 STR markers in common between 14 populations*

	PERU	Mexico	Ecuador	Bolivia	Hispano america	Aymaras Peru	Ashaninkas Peru	Loreto Peru	Amazonas Peru	Norte Peru	Sur Peru	Centro Peru	Oriente Peru	Lima_ Callao Peru
PERU	0,0000													
Mexico	0,0114	0,0000												
Ecuador	0,0037	0,00929	0,0000											
Bolivia	0,0029	0,01717	0,00841	0,0000										
Hispanoamerica	0,0162	0,01519	0,01039	0,02387	0,0000									
Aymaras_Peru	0,0059	0,0225	0,01388	0,00136	0,03259	0,0000								
Ashaninkas_ Peru	0,0264	0,03146	0,02342	0,03392	0,04442	0,03942	0,0000							
Loreto_Peru	0,0110	0,0201	0,01265	0,0153	0,03187	0,01688	0,01663	0,0000						
Amazonas_Peru	0,0358	0,03947	0,03565	0,0411	0,04454	0,04311	0,04812	0,02691	0,0000					
Norte_Peru	0,0002	0,01139	0,00252	0,00384	0,01567	0,00733	0,02418	0,00953	0,03345	0,0000				
Sur_Peru	0,0010	0,01402	0,0055	0,00058	0,02101	0,00266	0,02816	0,00988	0,03403	0,00153	0,0000			
Centro_Peru	0,0009	0,01595	0,00709	0,00235	0,02264	0,00459	0,03114	0,01332	0,03808	0,00165	0,00099	0,0000		
Oriente_Peru	0,0018	0,01111	0,00359	0,00739	0,01699	0,01095	0,01807	0,00446	0,02786	0,00098	0,00319	0,00368	0,0000	
Lima_Callao_ Peru	0,0001	0,01232	0,0042	0,0021	0,01648	0,00555	0,02888	0,01284	0,03799	0,00064	0,00126	0,00081	0,0027	0,0000

Table 3. *Average Fst values obtained between pairs of populations for 15 STR markers in common between 12 peruvian populations*

	Ashanincas Peru	Loreto Peru	Amazonas Peru	Madre de Dios Peru	Norte Peru	Sur Peru	Centro Peru	Oriente Peru	Lima_ Callao Peru	COASTAL PERU	SIERRA PERU	JUNGLE PERU
Ashanincas_Peru	0.00000											
Loreto_Peru	0.01663	0.00000										
Amazonas_Peru	0.04812	0.02691	0.00000									
Madre_de_Dios_ Peru	0.11652	0.08911	0.12808	0.00000								
Norte_Peru	0.02418	0.00953	0.03345	0.08373	0.00000							
Sur_Peru	0.02816	0.00988	0.03403	0.08066	0.00153	0.00000						
Centro_Peru	0.03114	0.01332	0.03808	0.07857	0.00165	0.00099	0.00000					
Oriente_Peru	0.01807	0.00446	0.02786	0.08503	0.00098	0.00319	0.00368	0.00000				
Lima_Callao_ Peru	0.02888	0.01284	0.03799	0.07552	0.00064	0.00126	0.00081	0.00270	0.00000			
COASTAL_PERU	0.02732	0.01143	0.03554	0.08300	0.00011	0.00222	0.00195	0.00153	0.00025	0.00000		
SIERRA_PERU	0.02825	0.01292	0.03869	0.08301	0.00156	0.00026	0.00017	0.00409	0.00076	0.00206	0.00000	
JUNGLE_PERU	0.02116	0.00565	0.03365	0.09034	0.00099	0.00069	0.00225	0.00139	0.00188	0.00013	0.00202	0.00000

Comparisons of the individual population samples from the coastal, sierra and jungle regions of Peru with the population of the macroregions of northern, central, southern, eastern Peru, Lima – Callao [11] and the population of the jungle of Peru with the population of Loreto [6] generated Fst vales <0.01. The comparisons of the coastal, sierra and jungle regions of Peru with the populations of the Amazonas, Madre de Dios [6] and the Ashaninka indigenous people [12] and only the populations of the coastal and highland regions of Peru with the population of Loreto [6] generated Fst values >0.01 (Figure 1 and Table 3).

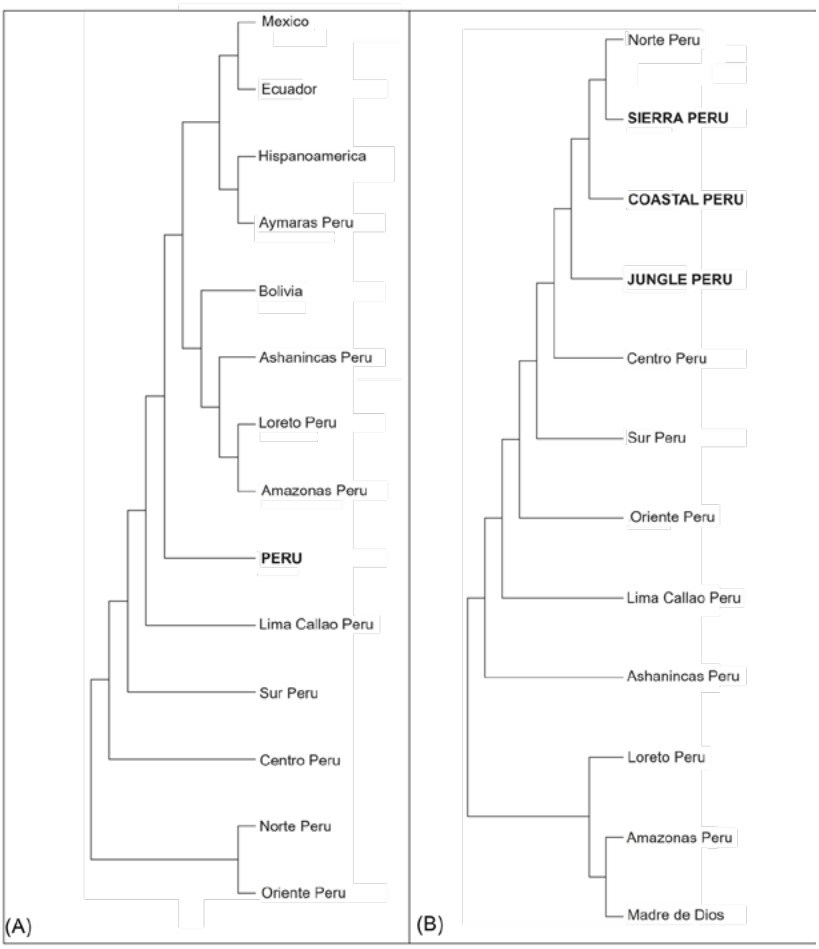


Figure 1. UPGMA tree based on 15 STR markers in common among the fourteen populations (A) and the twelve peruvian populations (B) and on Reynolds distances

Conclusion

Little or no detectable population genetic subdivision was observed with the combined Peruvian population compared with Ecuador, Bolivia, the Aymara people and the macroregions of Peru. However, there are differences with the population sample of Peruvian mestizo compared and population samples from Mexico, Hispanic America, the Ashaninka people, and the departments of Loreto and Amazonas in Peru. When comparing the populations of the coast, sierra and jungle of Peru with other Peruvian populations, there are differences with the population samples of the departments of Amazonas, Madre de Dios, and the Ashaninkas and only the populations of the coast and sierra of Peru with the population of Loreto. A database of allele frequencies for 23 autosomal STR markers further describing the Peruvian population(s) is presented.

Acknowledgments

This present research was supported by the Dirección de Criminalistica of the Policia Nacional del Perú and the Faculty of Biology of the Universidad Nacional Mayor de San Marcos.

Conflict of interest statement

None

References

[1] D.N. Harris, W. Song, A.C. Shetty, et al., Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. Proc Natl Acad Sci U S A. 115 (2018) E6526-E6535.

[2] E. Delgado, C.D. Neyra, Allele frequencies of 21 autosomal STR markers in a mixed race Peruvian population applied to forensic practice, Rev. Espanola de Medicina Leg. 45,3 (2018), 92-97. https://doi.org/10.1016/j.reml.2018.09.002

[3] C.D. Neyra, E. Delgado, C. Robles, et al., Population genetic study of a Peruvian population using human identification STRs, Int J Legal Med. 134,6 (2020), 2071-2073. https://doi.org/10.1007/s00414-020-02418-6

[4] C.D. Neyra, M. Suarez, E. Cueva, et al., Identificación genética de recién nacidos en Perú: un estudio piloto. Rev Chil Pediatr 90,1 (2019), 26-35. https://dx.doi.org/10.32641/rchped.v90i1.730

[5] C.D. Neyra-Rivera, B. Budowle, Genetic variation of different Peruvian populations using 23 autosomal STR markers. Forensic Sci Int Genet Suppl Ser 8. (2022), 259–262. https://doi.org/10.1016/j.fsigss.2022.10.055

[6] C.D. Neyra-Rivera, E. Delgado, F. Diaz, et al., Genetic study with autosomal STR markers in people of the Peruvian jungle for human identification purposes, Can Soc Forensic Sci. 54, 3 (2021), 117-138. https://doi.org/10.1080/00085030.2021.1933811

[7] C.D. Neyra-Rivera, C.S. Robles, E. Delgado, et al., Analysis of 27 STR markers of the chromosome Y in the population of the jungle from Peru, Rev. Espanola de Medicina Leg. 0377-4732 (2023). https://doi.org/10.1016/j.reml.2022.10.001

[8] C.D. Neyra-Rivera, A. Ticona, E. Delgado, et al., Allelic frequencies with 23 autosomic STRS in the Aymara population of Peru, Int J Legal Med. 135, 3 (2020):779-781. https://doi.org/10.1007/s00414-020-02448-0

[9] C.D. Neyra-Rivera, A. Ticona, E Delgado, et al., Population data of 27 Y-chromosome STRS in Aymara population from Peru, Aust J Forensic Sci. (2021)1-15. https://doi.org/10.1080/00450618.2021.1882571

[10] C.D. Neyra-Rivera, E.D. Ramos, E.G. Ingunza, et al., Analysis of 27 Y-chromosomal STR loci of the Mestizo Peruvian population, Egypt J Forensic Sci 13, 47 (2023). https://doi.org/10.1186/s41935-023-00366-x

[11] M.E. Bermejo, J. Ge, B. Budowle, C. Infante, C.D. Neyra-Rivera, Genetic study with 21 autosomal STRs in five Peruvian macro regions for human identification purposes, Legal Med 57 (2022), 1344-6223. https://doi.org/10.1016/j.legalmed.2022.102073

[12] D.H.Tineo DH (2017), Caracterización genética de una muestra poblacional ashaninka en el distrito de Puerto Bermúdez, Pasco- Perú, empleando marcadores STR autosómico y del cromosoma Y [Tesis de Maestría]. Universidad Nacional Mayor de SanMarcos, 2017.

[13] A. Tereba, Tools for analysis of population statistics. Profiles DNA. (1999) 14-16.

[14] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinform. (2005), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2658868/

[15]. K. Tamura, G. Stecher,D. Peterson, et al., MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol. 30 (2013) 2725-2729.

[16] R. Flores-Espinoza, A.N. Angulo-Pozo, A. Garzón-Salazar, et al., Evaluating population structure of Ecuador for forensic STR markers, Forensic Sci Int Genet Suppl Series 8 (2022) 102-104. https://doi.org/10.1016/j.fsigss.2022.09.036.

[17] O. Rocabado, O. García, I. Yurrebaso I, et al., (2009) Population genetic data for 15 STR loci (Identifiler™ kit) in Bolivia, Legal Med. 11, 6 (2009) 302–304. https://doi.org/10.1016/j.legalmed.2009.07.002.

[18] J.A. Aguilar, G. Locia, B. Lopez, et al., Forensic parameters and admixture in seven geographical regions of the Guerrero state (South, Mexico) based on STRs of the Globalfiler kitAnn. Hum. Biol. 45 (2018) 524-530. https://doi.org/10.1080/03014460.2019.1568548.

[19] Hill C, Duewer D, Kline M, et al., US population data for 29 autosomal STR loci. Forensic Sci Int Genet 7, 3 (2013) e82–e83. doi.org/10.1016/j.fsigen.2012.12.004.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 571-577
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Genetic Analysis based on 23 autosomal STRs of the Moroccan population

H. El Ossmani[1,4], A. Bouzouggar[4], K. Cheffi[2], A. El Khair[2], N. Dahbi[2], A. Hilali[2], B. Gazzaz[3]

[1]*Forensic Institute of Gendarmerie Royale, Rabat Rabat, Morocco*
[2]*Hassan First University of Settat, Higher Institute of Health Sciences, Laboratory of Health Sciences and Technologies, Settat, Morocco, Settat Setatt, Morocco*
[3]*Institut des Analyses Génétiques de la Gendarmerie Royale; Rabat; Morocco, Rabat Rabat, Morocco*
[4]*Institut National des Sciences de l'Archéologieet du Patrimoine; Rabat; Morocco, Rabat Rabat, Morocco*

Corresponding author: Hicham El Ossmani
Angle Avenue Ben Barka, Rue Ghomara Souissi Rabat Morocco,h.elossmani@gr.ma.

Abstract

Morocco has long been a subject of interest in genetic research because of its high heterogeneity manifested through various ethnic groups (Arabs, Berbers, and Sahraouis). This study describes, for the first time, the genetic profile of the mixed population located in center of Morocco. Twenty three (23) STR was explored on a random sample of **750 unrelated** healthy individuals from different regions in Morocco. Based on Nei genetic distance, a Neighbor-Joining tree and a Principal coordinate analysis (PCoA) were implemented. For a more detailed approach, a locus by locus pairwise genetic distance (Fst) was calculated. After Bonferroni's correction, only two loci (TH01 and D18S51)deviated from Hardy–Weinberg equilibrium. The Moroccan population was clustered with Northwest African populations. It was the closest to the Berber population of Azrou. It shared close genetic affinities with all North African populations, with a west to east gradient emphasizing varying degrees of Berber or Arab ancestry contributions, as well as with Sub-Saharan populations and Iran. The genetic affinity with Arabian populations was weak and no clear evidence of significant genetic contribution from European population.The Moroccan population seems to embody major North African characteristics, strongly related to a Berber ancestry, with a relatively minor influence from an Arabian origin, as well as a close genetic affinity with Sub-Saharan populations and, surprisingly, with Iran.

Keywords
STRs, Population data, Morocco, Arabs, berbers, sahraouis.

1. Introduction

STRs (Short tandem repeats) are widespread in the genomes of all living organisms, from bacteria to humans [1; 2]. Many scientists have studied the history of human evolution and migration using STR loci [3; 4].Because of their high level of polymorphism. In Morocco, multiple studies based on STRs were conducted to study the genetic structure of this population [5; 6]. Morocco is one of the ancient countries whose history goes back to the mists of time.It has attracted the attention of different civilizations (Phoenicians, Romans, Vandals, Arabs, Portuguese, Spain and French) because of its strategic geographical location and its wealth. The documented history of Morocco begins with the Phoenician civilization that colonized the coasts of Morocco between the 8th and 6th centuries BC [7]. Through some fossils and the remains of Homo sapiens, found in Jebel Irhoud (western Morocco), the origin of Homo sapiens goes back more than 315,000 years instead of 200,000 years ago as recognized by the scientific community [8].

2. Material and methods

2.1. Sample collection and DNA preparation

Buccal swabs were collected from 750 randomly selected, unrelated healthy individuals of both genders from different regions in Morocco with ancestry traced back at least three generations. All samples were obtained with the participants' informed consent. DNA was extracted using Chelex® 100 (Bio-Rad Laboratories, USA) method [9].

2.2. PCR amplification and genotyping by capillary electrophoresis

Multiplexed PCR amplification of 23 short tandem repeat (STR) loci (D3S1358,VWA, CSF1PO, D16S539, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, THO1, FGA, D22S1045, D13S317, D5S818, D7S820, SE22, D2S1338, D10S1248, D1S1656, D12S391, Y indel and DYS391) was performed using the AmpFlSTR® Globalfiler® kit in PCR System 9700 ABI 96-Well Thermal Cycler according to the manufacturer's specifications. Electrophoresis and typing were

carried out in the ABI 3130xl genetic analyzer. Genotypes were assigned using the GeneMapper® ID-X v1.1 software by comparison with reference allelic ladders.

2. 3 Phylogenetic tree

The phylogenetic tree (UPGMA) was constructed based on DA genetic distance [10].These calculations were performed using the web version of POPTREE [11].

3. Results

3.1. STRs Markers characterization

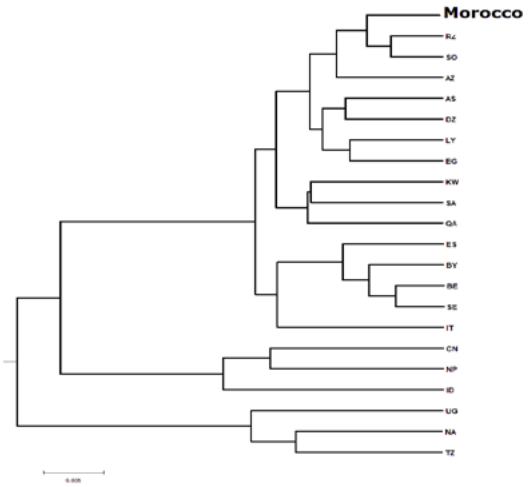


Figure 1: UPGMA tree showing the phylogenetic relationships of the Moroccanpopulation among 21 worldwide populations.

Populations studied: Morocco (present study);AS: Berber-speaking from Asni (High Atlas Mountains of Morocco);RZ: Arabic-speaking population of Rabat-Sale-Zemmour-Zaer;AZ: Berber-speaking population of Azrou;SO: Southern Morocco population;DZ: Berber population of Bejaia (Algeria);LY: Libya; EG: Egypt; UG: Karamoja population (Uganda);NA: Ovambo population (Namibia);TZ: Tanzania;KW: Kuwait;QA: Qatar;SA: Central Region of Saudi Arabia;ID: Indonesia;CN: China Han in Jilin Province (China);NP: Tibetan population (Nepal);ES: Andalusians from Huelva (Spain);BY: Belarus;BE: Belgium;IT: Sardinia island (Italy);SE: Sweden.

3.2Principal component analysis (PCA)

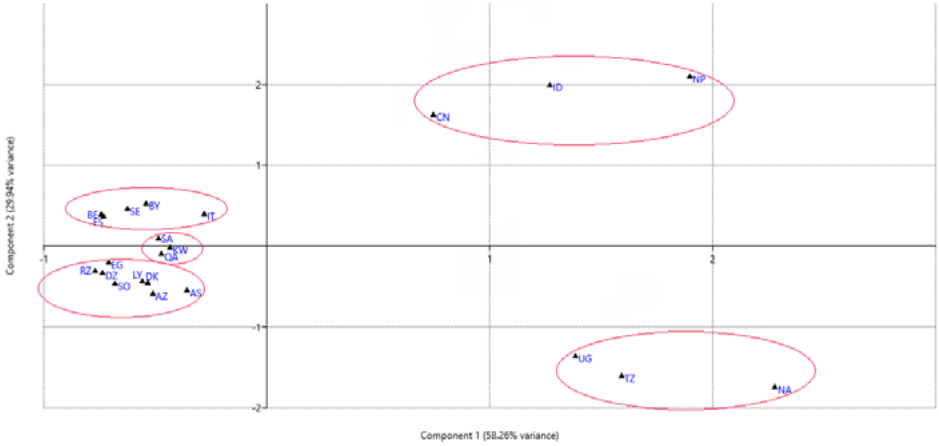


Figure 2: Principal components analysis (PCA) based on Nei's standard genetic distances at the 23 autosomal STR loci between the Moroccan population and 21worldwide populations.

Populations studied: Morocco (present study); AS: Berber-speaking from Asni (High Atlas Mountains of Morocco);RZ: Arabic-speaking population of Rabat-Sale-Zemmour-Zaer;AZ: Berber-speaking population of Azrou;SO: Southern Morocco population;DZ: Berber population of Bejaia (Algeria);LY: Libya; EG: Egypt; UG: Karamoja population (Uganda);NA: Ovambo population (Namibia);TZ: Tanzania;KW: Kuwait;QA: Qatar;SA: Central Region of Saudi Arabia;ID: Indonesia;CN: China Han in Jilin Province (China);NP: Tibetan population (Nepal);ES: Andalusians from Huelva (Spain);BY: Belarus;BE: Belgium;IT: Sardinia island (Italy);SE: Sweden.

4. Discussion

A comparison of genetic distance indices was made by using Arlequin v3.5.2.2 software, based on the P-values of Wright's F-statistic (fixation index - FST) between (Morocco) and 21 worldwide populations. The tests revealed that no significant difference ($P<0.05$) was observed between the study population and the North African populations (Arabic-speaking population of Rabat-Sale-Zemmour-Zaer, Southern Morocco population, Berber-speaking population of Azrou, Berber-speaking from Asni from High Atlas Mountains of Morocco, Berber population of Bejaia from

Algeria, Libyan population) at all loci; except for the Egyptian population, the difference was detected in 2 out of 23 STR loci. This difference seems to be due to the geographical distance between Egypt population (Northeast Africa) and the study population (Northwest Africa). In fact, geographically close populations are also genetically similar and when the geographical distance increases, the genetic diversity between groups becomes more important [10]. Whereas a significant differences were found between Morocco and Andalusians from Huelva (Spain) at 3 loci; Sardinia island (Italy) at 3 loci; Qatar population at 4 loci; Kuwait population at 5 loci; Saudi Arabia population at 5 loci; Sweden population at 5 loci; Belgium population at 6 loci; Belarus population at 6 loci; Karamoja population (Uganda) at 9 loci; Indonesia population at 9 loci; China Han in Jilin population at 11 loci; Ovambo population (Namibia) at 12 loci; Tanzania population at 12 loci; Tibetan population (Nepal) at 13 loci.

The figure 1 shows the result of the phylogenetic tree (UPGMA) performed after the allelic frequencies of eight North African populations, three Sub-Saharan populations, three Arabian Peninsula populations, five European populations and three East Asian populations were included in the analysis as part of a global study. As could be seen from the phylogenetic tree (Figure), the Moroccan population is grouped with Arabic-speaking population of Rabat-Sale-Zemmour-Zaer (Morocco), Southern Morocco population and Berber-speaking population of Azrou (Morocco), Berber-speaking from Asni in High Atlas Mountains (Morocco), Berber population of Bejaia (Algeria), Libya, Egypt; and appears to form a sister cluster to the group containing the Arabian Peninsula populations (Qatar, Central Region of Saudi Arabia and Kuwait).

Principal component analysis (PCA) (Figure 2) revealed that the North African populations appear to be heterogeneous compared to the other populations included in this analysis, which reflects an important genetic affinity expressing at the same time the geographical, socio-cultural and historical closeness. The Moroccan population was grouped with North African populations (Arabic-speaking population of Rabat-Sale-Zemmour-Zaer in Morocco, Southern Morocco population, Berber-speaking population of Azrou in Morocco, Berber population of Bejaia in Algeria, Berber-speaking from Asni in High Atlas Mountains of Morocco and Egyptian population). However, the populations of the Arabian Peninsula (Kuwait, Qatar and Central Region of Saudi Arabia) are not far from the North African populations. Andalusians from Huelva (Spain) and Belgium form a group with Sardinia Island (Italy), Sweden and Belarus in the upper left quadrant, which shows a great similarity among the European populations. While the East Asian countries

(China Han in Jilin province, Indonesia and Tibetan population) cluster in the upper right. The populations of Sub-Saharan Africa (Uganda, Tanzania and Namibia) are grouped in the lower right quadrant.

5. Conclusion

The phylogenetic tree, PCA plots and population comparison tests revealed a significant affinity between the Morocco population and North African populations. **In fact, the 23 STRs loci in this study were highly informative**, thus proving their usefulness for forensics and population genetic studies.

Conflict of intereststatement

The authors declare that they have no conflict of interest.

Ethical Approval

This study was approved by the Biomedical Research Ethics Committee (CERBC) of Casablanca, Morocco. The Ethics Committee is based on the Declaration of Helsinki 2008.

Funding

No funds, grants, or other support was received.

Author's information

Hicham El Ossmani: https://orcid.org/0000-0002-1592-6733

References

[1] EP. Rocha, A. Danchin, A. Viari. Functional and evolutionary roles of long repeats in prokaryotes. Res Microbiol. Nov-Dec;150 (9-10) (1999) 725-33.

[2] MV. Katti, PK. Ranjekar, VS. Gupta. Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol Biol Evol. Jul;18(7)(2001) 1161-7.

[3] NA. Rosenberg, JK. Pritchard, JL. Weber et al, Genetic structure of human populations. Science, 298 (5602), (2002) 2381-2385.

[4] DB, Goldstein, A. Ruiz Linares, LL, Cavalli-Sforza et al. Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci U S A. Jul 18;92(15): (1995) 6723-7.

[5] C.Coudray, E.Guitard, C. Keyser-Tracqui et al. Population genetic data of 15 tetrameric short tandem repeats (STRs) in Berbers from Morocco. Forensic Science International, 167 (1) (2007) 81–86.

[6] H. El Ossmani, B.Bouchrif,R Aboukhalid, Bouabdillah et al. Assessment of phylogenetic structure of Berber-speaking population of Azrou using 15 STRs of Identifiler kit. Legal Medicine, 12(1) (2010) 52–56.

[7] C. R. Pennell Morocco: From Empire to Independence, page 5. 2003.

[8] Callaway, E. Oldest Homo sapiens fossil claim rewrites our species' history. Nature (2017).

[9] PS. Walsh, DA. Metzger andR. Higuchi. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques. 1991 Apr; 10(4):506–13.

[10] M. Nei, F. Tajima and Y. Tateno. Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol 19(1983)153-170.

[11] N.Takezakim M Nei and Tamura K. POPTREEW: Web version of POPTREE for constructing population trees from allele frequency data and computing some other quantities. Molecular Biology and Evolution, (2014).

[12] F Calafell, A. Perez-Lezaun, andJ. Bertranpetit, Genetic distances and microsatellite diversification in humans. Hum Genet. 106 (1) (2000) 133-134.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 579-586
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Genetic variation in the Chanca population using 23 autosomal STR markers

Edgardo Delgado Ramos[1], Carlos Neyra-Rivera[2*], Sonia Guillen Ramirez[3], Bruce Budowle[4]

[1]*Instituto de Medicina Legal y Ciencias Forenses, Lima, Perú*

[2]*Instituto de Investigaciones en Ciencias Biomédicas, Universidad Ricardo Palma, Lima, Perú*

[3]*Laboratorio de Microbiología Molecular y Biotecnología, Facultad de Ciencias Biológicas, Universidad Nacional Mayor de San Marcos, Lima,Peru.*

[4]*Department of Forensic Medicine, University of Helsinki, Finland*

** Corresponding author at: Av. Alfredo Benavides 5440, Santiago de Surco, Lima 15039, carlosdavidmp@outlook.es*

Abstract

Genetic variation was determined for 23 autosomal STR markers in the Chanca population residing in the province of Abancay-Peru. The STR profiles of 150 unrelated Chancas individuals from Chincheros (n= 93), Andahuaylas (n= 49), Abancay (n=6), Antabamba (n= 1) and Aymaraes (n= 1) were generated and analyzed. The Arlequin v3.5.2.2 software was used to determine the genetic distances (Fst) of the studied populations. Little or no detectable population genetic subdivision was observed between the population sample of Peruvian mestizo compared with populations from Mexico, Ecuador, Bolivia, Aymara from Peru, North from Peru, South from Peru, Center from Peru, East from Peru and Lima-Callao from Peru.

Keywords

Autosomal STR, Chancas, human identification, genetic distance.

Introduction

The genetic heritage of native populations of South America has been affected by the Inca and Spanish conquests, and studies with various genetic markers show

substantial evidence of population substructure in different Peruvian populations [1]. Additionally, multicultural variety of Peru can impact genetic variation of Peruvians and/or isolated populations [2]. Based on genomic analyses, Native Americans in Peru show greater affinities with the Andean indigenous peoples, both Quechua and Aymara [3]. Despite the growing number of studies on different Peruvian populations [4-13], there still are populations for which there are no genetic data. One such population is the Chanca, which is one of the most notable in ancient Peru. The Chanca originate from the Quechuas and were divided into Anansaya and Urinsaya. In an attempt to annex other ethnic groups, they were defeated by the Incas of Cusco in the decade of 1430. The population continued as a wamani of the Tahuantinsuyu and then as a corregimiento of the viceroyalty, persisting until today with the category of province of the department and region of Apurímac [14]. Due to the historical importance of the Chanca, the present research aimed to analyze the genetic variation of 23 autosomal STR markers in the Chanca population residing in Peru.

Material studied, methods, techniques

This study was approved by the Ethics Committee N°042-2023-CBE-FCB-UNMSM of the Faculty of Biological Sciences of the Universidad Nacional Mayor de San Marcos (Peru), and all participants freely and voluntarily signed their informed consent. Blood samples were collected from 150 Chancas residing in the provinces of Chincheros (n= 93), Andahuaylas (n= 49), Abancay (n= 6), Antabamba (n= 1) and Aymaraes (n= 1) in the department of Apurimac and stored on FTA cards. DNA was extracted from five 1.2mm punches using the Chelex-100 method and quantified using a Qubit fluorometer (Thermo Fisher Scientific) according to the manufacturer's instructions. The genetic markers were amplified by PCR multiplex using the Verifiler Kit (Thermo Fisher Scientific), with 2 ng input DNA per sample in a reaction volume of 25 µL. The amplified products were subjected to capillary electrophoresis on the ABI 3500XL, and the results were analyzed with the GeneMapper® ID-X v1.5 software (Thermo Fisher Scientific). Discrimination power (PD), polymorphic information content (PIC), exclusion probability (PE), and observed heterozygosity (Ho) were calculated with the Power Stats V1.2 software [15]. The Arlequin v3.5.2.2 software was used to determine the genetic distances (Fst) of the studied populations [16]. The genetic distances were used to generate a UPGMA tree using the MEGA 6.06 program [17].

Results and Discussion

In the Chanca population sample all markers had PIC values ≥0.50 with the highest PIC values observed for the markers D6S1043 (0.8385), D18S51 (0.8471) and PENTA E (0.8795). The highest PD values were observed in the markers D6S1043 (0.9567), D18S51 (0.9622) and PENTA E (0.9732). The most informative marker for PE was FGA (0.7413) and the lowest was TPOX (0.2601). Only one marker, D6S1043 (0.0431), did not meet Hardy-Weinberg expectations but after applying the Bonferroni correction (significance value p=0.0022) this maker was no longer considered significantly departing from expectations. (Table 1).

The pairwise population comparisons with the Chanca that resulted in an Fst <0.01 were the populations of Mexico [18], Ecuador [19], Bolivia [20], Aymara from Peru [10], Noth from Peru, South from Peru, Center from Peru, East from Peru, and Lima-Callao from Peru [13]. The Ashanincas [21], Loreto and Amazonas (peruvians populations) [8] comparisons resulted in Fst values >0.01, which is indicative of notable population substructure (Table 2). In the UPGMA tree (Figure 1) the Chanca population is close to the Aymara population [10], while the Loreto population is close to the Amazonas population [8] and both were well separated from the Ashaninka population [21] (Figure 1).

Table 1. Genetic parameters of the Chancas sample population

Genetic parameters	D3S1358	vWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	D10S1248	D1S1656	D12S391	D2S1338	PENTA E	D6S1043	PENTA D
PD	0,7996	0,8671	0,9170	0,7996	0,7690	0,9327	0,9454	0,9622	0,7486	0,9392	0,7969	0,9564	0,7764	0,7764	0,9082	0,8703	0,8428	0,9505	0,9170	0,9332	0,9732	0,9567	0,9259
PIC	0,5824	0,6580	0,7507	0,5824	0,5296	0,7782	0,8151	0,8471	0,5000	0,7946	0,5689	0,8378	0,5360	0,5360	0,7283	0,6688	0,6344	0,8277	0,7552	0,7743	0,8795	0,8385	0,7667
PE	0,3882	0,4386	0,6115	0,3882	0,2601	0,6115	0,7147	0,7014	0,2752	0,6240	0,2890	0,7413	0,3072	0,3072	0,4928	0,4283	0,4599	0,7147	0,6240	0,5867	0,7280	0,6240	0,5990
Het	0,6733	0,7067	0,8067	0,6733	0,5733	0,8067	0,8600	0,8533	0,5867	0,8133	0,6067	0,8733	0,6133	0,6133	0,7400	0,7000	0,7200	0,8600	0,8133	0,7933	0,8667	0,8133	0,8000
P-Value	0,6166	0,3591	0,7570	0,3030	0,5000	0,4187	0,6953	0,6329	0,3547	0,6707	0,3477	0,8237	0,9003	0,3576	0,4320	0,8237	0,8577	0,3276	0,3750	0,7061	0,2399	0,0431	0,6120
PM	0,2004	0,1329	0,0830	0,2004	0,2310	0,0673	0,0546	0,0378	0,2514	0,0608	0,2031	0,0436	0,2236	0,2236	0,4928	0,1297	0,1572	0,0495	0,0830	0,0668	0,0268	0,0433	0,0741
Het(exp)	0,6456	0,7103	0,7872	0,6776	0,5955	0,8076	0,8379	0,8651	0,5449	0,8190	0,6234	0,8569	0,6151	0,6216	0,7582	0,7203	0,6918	0,8487	0,7872	0,7983	0,8922	0,8555	0,7965
Het(obs)	0,6733	0,7067	0,8067	0,6933	0,5733	0,8067	0,8600	0,8533	0,5867	0,8200	0,6067	0,8733	0,6133	0,6200	0,7400	0,7000	0,7200	0,8600	0,8133	0,7933	0,8667	0,8133	0,8000
N	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150

Table 2. *Average Fst values obtained between pairs of populations for 15 STR markers in common between 13 populations*

	Mexico	Ecuador	Bolivia	CHANCAS PERU	Aymaras Peru	Ashanincas Peru	Loreto Peru	Amazonas Peru	Norte PeruSur	Peru	Centro Peru	Oriente Peru	Lima_ CallaoPeru
Mexico	0.00000												
Ecuador	0.00929	0.00000											
Bolivia	0.01717	0.00841	0.00000										
CHANCAS_ PERU	0.01609	0.00716	0.00270	0.00000									
Aymaras_ Peru	0.02250	0.01388	0.00136	0.00490	0.00000								
Ashanincas_ Peru	0.03146	0.02342	0.03392	0.03269	0.03942	0.00000							
Loreto_Peru	0.02010	0.01265	0.01530	0.01466	0.01688	0.01663	0.00000						
Amazonas_ Peru	0.03947	0.03565	0.04110	0.04069	0.04311	0.04812	0.02691	0.00000					
Norte_Peru	0.01139	0.00252	0.00384	0.00167	0.00733	0.02418	0.00953	0.03345	0.00000				
Sur_Peru	0.01402	0.00550	0.00058	0.00140	0.00266	0.02816	0.00988	0.03403	0.00153	0.00000			
Centro_Peru	0.01595	0.00709	0.00235	0.00027	0.00459	0.03114	0.01332	0.03808	0.00165	0.00099	0.00000		
Oriente_Peru	0.01111	0.00359	0.00739	0.00546	0.01095	0.01807	0.00446	0.02786	0.00098	0.00319	0.00368	0.00000	
Lima_Callao_ Peru	0.01232	0.00420	0.00210	0.00130	0.00555	0.02888	0.01284	0.03799	0.00064	0.00126	0.00081	0.00270	0.00000

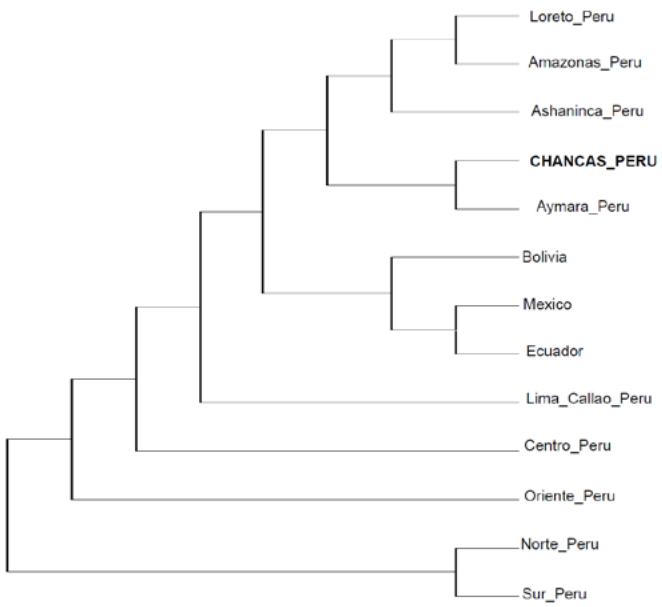


Figure 1. UPGMA tree based on 15 STR markers in common among the thirteen populations and on Reynolds distances

Conclusion

This study provides STR population data from the Chanca population, one of the ethnic groups of Peru, distinct from the mestizo populations of Peru (North, South, Center, East and Lima-Callao of Peru) and closer to the native populations (Ashanincas, Loreto and Amazonas of Peru). The results provide additional population data for statistical calculations in human identity testing cases in Peru.

Acknowledgments

This present research was supported by the Instituto de Investigaciones de Ciencias Biomédicas of the Universidad Ricardo Palma and the Faculty of Biology of the Universidad Nacional Mayor de San Marcos.

Conflict of interest statement

None

References

[1] E.K. Guevara, J.U. Palo, S. Översti, et al., Genetic assessment reveals no population substructure and divergent regional and sex-specific istories in the Chachapoyas from northeast Peru, PloS one, 15, 12 (2020), e0244497. https://doi.org/10.1371/journal.pone.0244497

[2] D. N. Harris, W. Song, A. C. Shetty, et al., Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. Proc Natl Acad Sci, 115.28 (2018), E6526-E6535. https://doi.org/10.1073/pnas.1720798115

[3] J.R. Homburger, A. Moreno-Estrada, C.R. Gignoux, et al., Genomic Insights into the Ancestry and Demographic History of South America. PLoS genetics, 11.12 (2015), e1005602. https://doi.org/10.1371/journal.pgen.1005602

[4] Delgado E, Neyra CD, Allele frequencies of 21 autosomal STR markers in a mixed race Peruvian population applied to forensic practice. Rev. Espanola de Medicina Leg, 45, 3 (2018), 92-97. https://doi.org/10.1016/j.reml.2018.09.002

[5] C.D. Neyra, E. Delgado, C. Robles, et al., Population genetic study of a Peruvian population using human identification STRs, Int J Legal Med. 134,6 (2020), 2071-2073. https://doi.org/10.1007/s00414-020-02418-6

[6] C.D. Neyra, M. Suarez, E. Cueva, et al., Identificación genética de recién nacidos en Perú: un estudio piloto. Rev Chil Pediatr 90,1 (2019), 26-35. https://dx.doi.org/10.32641/rchped.v90i1.730

[7] C.D. Neyra-Rivera, B. Budowle, Genetic variation of different Peruvian populations using 23 autosomal STR markers. Forensic Sci Int Genet Suppl Ser 8. (2022), 259–262. https://doi.org/10.1016/j.fsigss.2022.10.055

[8] C.D. Neyra-Rivera, E. Delgado, F. Diaz, et al., Genetic study with autosomal STR markers in people of the Peruvian jungle for human identification purposes, Can Soc Forensic Sci. 54, 3 (2021), 117-138. https://doi.org/10.1080/00085030.2021.1933811

[9] C.D. Neyra-Rivera, C.S. Robles, E. Delgado, et al., Analysis of 27 STR markers of the chromosome Y in the population of the jungle from Peru, Rev. Espanola de Medicina Leg. 0377-4732 (2023). https://doi.org/10.1016/j.reml.2022.10.001

[10] C.D. Neyra-Rivera, A. Ticona, E. Delgado, et al., Allelic frequencies with 23 autosomic STRS in the Aymara population of Peru, Int J Legal Med. 135, 3 (2020):779-781. https://doi.org/10.1007/s00414-020-02448-0

[11] C.D. Neyra-Rivera, A. Ticona, E Delgado, et al., Population data of 27 Y-chromosome STRS in Aymara population from Peru, Aust J Forensic Sci. (2021)1-15. https://doi.org/10.1080/00450618.2021.1882571

[12] C.D. Neyra-Rivera, E.D. Ramos, E.G. Ingunza, et al., Analysis of 27 Y-chromosomal STR loci of the Mestizo Peruvian population, Egypt J Forensic Sci 13, 47 (2023). https://doi.org/10.1186/s41935-023-00366-x

[13] M.E. Bermejo, J. Ge, B. Budowle, C. Infante, C.D. Neyra-Rivera, Genetic study with 21 autosomal STRs in five Peruvian macro regions for human identification purposes, Legal Med 57 (2022), 1344-6223. https://doi.org/10.1016/j.legalmed.2022.102073

[14] W.E. Soriano. El reino chanca. Investigaciones sociales, 21.38 (2017), pp. 139-172. https://doi.org/10.15381/is.v21i38.14223

[15] Tereba, A. (1999). Tools for analysis of population statistics. Profiles DNA: 14-16.

[16] Excoffier L., Laval G., Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2658868/

[17] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol. 2013; 30: 2725-9.

[18] J.A. Aguilar, G. Locia, B. Lopez, et al., Forensic parameters and admixture in seven geographical regions of the Guerrero state (South, Mexico) based on STRs of the Globalfiler kitAnn. Hum. Biol. 45 (2018) 524-530. https://doi.org/10.1080/03014460.2019.1568548.

[19] R. Flores-Espinoza, A.N. Angulo-Pozo, A. Garzón-Salazar, et al., Evaluating population structure of Ecuador for forensic STR markers, Forensic Sci Int Genet Suppl Series 8 (2022) 102-104. https://doi.org/10.1016/j.fsigss.2022.09.036.

[20] O. Rocabado, O. García, I. Yurrebaso I, et al., (2009) Population genetic data for 15 STR loci (Identifiler™ kit) in Bolivia, Legal Med. 11, 6 (2009) 302–304. https://doi.org/10.1016/j.legalmed.2009.07.002

[21] D.H.Tineo DH (2017), Caracterización genética de una muestra poblacional ashaninka en el distrito de Puerto Bermúdez, Pasco- Perú, empleando marcadores STR autosómico y del cromosoma Y [Tesis de Maestría]. Universidad Nacional Mayor de SanMarcos, 2017.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 587-591
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Single nucleotide polymorphism–short tandem repeat analysis system for Japanese people

Tsukasa Ohuchi[1], Sohtaro Mimasaka[1]

[1]*Department of Forensic Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan.*

** Corresponding author at: 2-1 Seiryo-machi, Aoba-ku, Sendai 980-8575, Japan, mail: tsukasa.ouchi.e6@thoku.ac.jp, tel: +81-22-717-8110, fax: +81-22-717-8112.*

Abstract

In forensic practice, human identification is generally performed by capillary electrophoresis targeting short tandem repeats (STRs). As an alternative, a single nucleotide polymorphism (SNP)-STR system has been reported that uses an amplification refractory mutation system to identify STRs and adjacent SNPs simultaneously [1,2]. This system has better discriminatory power than the traditional system and advantageously does not require expensive massively parallel sequencing (MPS) equipment. We have tried to develop a specific SNP-STR system targeting nine loci (rs4347015-D1S1656, rs74640515-D2S441, rs149466976-D3S4529, rs12187142-D5S2800, rs25768-D5S818, rs16887642-D7S820, rs75219269-vWA, rs9546005-D13S317 and rs11642858-D16S539) for common SNPs in Japanese people [3].

Eight out of nine loci, except for rs149466976-D3S4529, were successfully genotyped although stutter alleles were detected at two loci (rs4347015-D1S1656 and rs25768-D5S818). For rs149466976-D3S4529, no mutant allele was detected, so that adjustments of primer concentrations and PCR conditions were required to establish this SNP-STR system. We plan to continue development to improve the accuracy of DNA typing.

Keywords

Human identification, Compound marker, SNP-STR.

Introduction

In forensic practice, DNA typing is mainly performed targeting highly polymorphic short tandem repeats (STRs). However, it is expected that the discrimination

ability can be improved by adding information on single nucleotide polymorphisms (SNPs) adjacent to STRs. In this study, we targeted nine STR loci and their adjacent SNPs based on DNA polymorphism information of Japanese [3] and attempted to construct a reaction system using the widely used capillary electrophoresis (CE) method [1,2].

2. Material studied, methods

2.1. Material

From a dataset of Japanese people [3], nine SNP-STR loci were selected as targets (Table 1). Two fluorophore labeled allele-specific SNP primers (wild-type and mutant) and one reverse (or forward) primer were designed according to a previous report [1]. DNA extracts from the blood of four volunteers with known SNP-STR types were used as test samples. This study was approved by the Ethics Committee of Tohoku University Graduate School of Medicine (No. 2020-1-902).

Table 1. Target loci in this study

STR Loci	Chromosome	rsID[a]	Reference allele	Alternative allele	Distance from STR region (bp)[b]
D1S1656	1	rs4847015	C	T	6
D2S441	2	rs74640515	G	A	-25
D3S4529[c]	3	rs149466976	G	A	10
D5S2800[c]	5	rs12187142	C	T	12
D5S818	5	rs25768	A	G	16
D7S820	7	rs16887642	G	A	9
vWA	12	rs75219269	A	G	-7
D13S317	13	rs9546005	A	T	1
D16S539	16	rs11642858	A	C	19

a) SNP ID assigned by NCBI (National Center for Biotechnology Information). b) Plus and minus signs show the upstream and downstream direction from each STR region, respectively. c) STR loci not included in the GlobalFiler™ PCR Amplification Kit (Thermo Fisher Scientific).

2.2. Methods

In this study, two primer sets were prepared (set_A and B) (Fig. 1). Multiplex PCR was performed in 20 µL reactions containing 10 µL Platinum Multiplex PCR Master Mix (Thermo Fisher Scientific), 2.4 µL GC enhancer, 0.6 µL primer mix

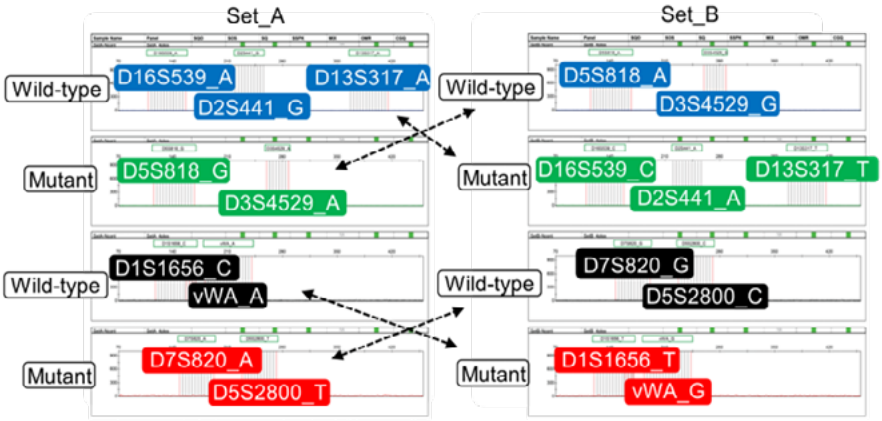


Fig. 1. Two primer sets (Set_A and B) were prepared

(one forward primer and one reverse primer, 3 µM of each), 5.4 µL nuclease-free water, and 1 µL genomic DNA (1 ng/µL). Thermocycling was performed in a Ploflex PCR System (Thermo Fisher Scientific) under the following conditions: initial denaturation at 95°C for 2 min; 30 cycles of 95 °C for 30 s, 58 °C for 90 s, 72 °C for 30 s; followed by a final extension step of 72 °C for 10 min. Electrophoresis was carried out using a genetic analyzer (SeqStudio, Thermo Fisher Scientific). Analysis was performed using Gene Mapper ID-X (version 1.6, Thermo Fisher Scientific).

Results

Eight out of nine loci, except for rs149466976-D3S4529, were successfully genotyped although stutter alleles were detected at two loci (rs4347015-D1S1656 and rs25768-D5S818) (Fig. 2a, b). For rs149466976-D3S4529, no mutant allele was detected (Fig. 2c).

Fig. 2. Representative results.

Discussion

As a result of attempting to develop a SNP-STR system using CE for Japanese, the following two issues were found out. Firstly, re-design of the primer for rs149466976-D3S4529 and adjustment of PCR conditions were needed to establish this SNP-STR system. Secondly, in addition to increasing the number of samples to ensure that typing is performed correctly, it is necessary to set appropriate stutter filter values.

Conclusion

This study demonstrated that eight of the nine SNP-STR loci commonly found in Japanese people can be identified by CE methods. It was suggested that this method could contribute to improving the accuracy of DNA typing without MPS. We plan to continue development this SNP-STR system to improve the accuracy of DNA typing.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP21K10515.

Conflict of interest statement

None.

References

[1] L Wang, P.M. Schneider, M.A. Rothschild et al., SNP–STR polymorphism: A sensitive compound marker for forensic genetic applications, Forensic Sci. Int. Genet. Suppl. Ser. 4 (2013) pp. e206-e207.

[2] H Jian, L Wang, M Lv, et al., A Novel SNP-STR System Based on a Capillary Electrophoresis Platform, Front Genet. 12 (2021) 636821.

[3] T Ohuchi, X Guan, E Hirai, et al., Allele frequencies of 31 autosomal short tandem repeat (auSTR) loci obtained using the Precision ID GlobalFiler™ NGS STR Panel v2 in 322 individuals from the Japanese population, Leg. Med. (Tokyo). 59 (2022) 102151.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 593-602
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Testing forensic efficacy of most recent available asSTRs and Y-STR markers in Indian population

Shivani Dixit[1,2*], Pankaj Shrivastava[3#], R.K. Kumawat[4] and Divya Shrivastava[2]

[1]*DNA Division, Central Forensic Science Laboratory, Chandigarh-160036, India*
[2]*Jaipur National University, Jaipur, Rajasthan, India*
[3]*DNA Division, Regional Forensic Science Laboratory, Jabalpur-482001 (MP), India*
[4]*DNA Division, State Forensic Science Laboratory, Jaipur, Rajasthan, India*

#Corresponding author - pankaj.shrivastava@rediffmail.com Phone: +91-9424371946

Abstract
Multiplex kits are made commercially available post publication of validation studies as per SWGDAM guidelines, for the purpose of forensic DNA examination. Such available multiplex kits are used in resolving criminal cases and also in population based genetic studies. Prior to their use these kits are not tested specifically on the populations where these kits are being used. In this study an attempt is made to test the most recent asSTRs and Y-STR multiplex systems on Indian population, to evaluate their forensic efficacy and to decipher best set of asSTR and Y-STR markers in Indian population. The study has shed light on what we coined to term as the Indian Standard DNA Marker System (ISDMS) that draws relevance for application on forensic DNA analysis, data basing and population genetic studies.

Keywords
Forensic, STRs, Indian Standard DNA Marker System (ISDMS), Database

Introduction

Short Tandem Repeat (STR) markers are well established and have historically been the gold standard for forensic DNA applications across the world [1]. Over the three and half decades, forensic DNA technology seen an exponential development with regards to the marker selection and the methods of analysis, fostered by an

ever-growing requirements in the development of competent, efficient, resources with high discriminative power and turn-around time [2], [3], [4].Various studies have reported application of diverse multiplex kits being used for forensic analysis[5], [6], [7]. Despite the much-needed validation of these multiplex kits, before their forensic DNA application, they should also be tested specifically on the populations on which these kits are being used, this would generate more meaningful scientific interpretations of the experimental outcomes[8].

It is highly welcomed that the kits have to be validated under the SWGDAM guidelines and manufactured under forensic grade (such as ISO 18385), which ensures confident use of these solutions in forensic casework. Combined DNA Index System (CODIS), Core Loci Working Group, the European Network of Forensic Science Institutes (ENFSI), and the European DNA Profiling Group (EDNAP), *National DNA Index System (NDIS), European Standard Set (ESS) etc., are* recommended set of polymorphic STR loci for the USA, European population for forensic DNA application and DNA data basing purposes. However, the above-mentioned standard set of loci should not be considered as mandatory for application on Indian population forensic sample analysis, and have been prompted by various studies to be not completely sufficient in case work analysis. As it is well documented that Indian population with over 1.5 billion people is diverse and limiting DNA analysis only to a defined set of CODIS/EDNAP/NDIS/ESS based markers might not be sufficient. Thus, it is highly important and is the need of hour to develop an Indian Standard DNA Marker System (ISDMS) which includes the best informative set of STR markers that cater to effective DNA analysis as well as data basing applications suiting the Indian scenario.

This report is prompted by the vast diversity in the genetic structure of Indian populations, with 4693 documented population groups that include 2205 major communities, 589 segmented units and 1900 territorial units India is a world in itself [9], [10]. Despite this wide range of genetic diversity in Indian population, till date there are no specific set of polymorphic loci that have be determined to define Indian population. This gap, drew our attention to be find best set of polymorphic loci for the Indian population, present study was conducted using all the recently used new generation STR based multiplex systems available in India.

Material studied, methods, techniques

Participants, Sample collection and genotyping

Randomly selected one thousand nine hundred and seventy (1970) individuals belonging to highly diverse Indian population were selected for sample collection. Volunteers' self-declared and written informed consent were obtained and peripheral blood samples were collected in EDTA blood vial following the Declaration of Helsinki, [11].

Genomic DNA was extracted by, using organic, automated approaches and direct amplification. The quantity and quality of the extracted DNA was checked by RT-PCR methods of PowerQuantTM (Promega Inc,USA) and Quantifiler®Trio quantification kit (Thermo Fisher Scientific, USA) as per manufacturer's protocol.

The Studied new generation autosomal STR loci markers were amplified via polymerase chain reaction (PCR) using multiple STR amplification systems, namely, PowerPlex Fusion 6C™ (374 samples)[12], VersaPlex™ 27PY (147 samples) [13], GlobalFiler™ (508 samples)[14], Verifiler Plus™(200 samples)[15], PowerPlex Y23 (202 Samples), Yfiler™ Plus (539 samples), on Veriti™ and ABI9700™ thermal cyclers (Thermo Fisher Scientific, USA) and as per the manufacturer's recommended protocol. The amplicons were separated using capillary electrophoresis on 3500XL™ Genetic Analyzer (Thermo Fisher Scientific, USA). The Data was analyzed using Gene Mapper ID-X™ versions1.5 and 1.6 software (Thermo Fisher Scientific, USA) as per developers 'settings.

ANALYSIS OF DATA: Allele frequency of the studied STR loci was calculated by using GenAlEx 6.5 software[16]. Forensic parameters viz., polymorphic information content (PIC), the power of discrimination (PD), the power of exclusion (PE), matching probability (P_m) and paternity index (PI),Observed heterozygosity (H_{obs}),were calculated by using the FORSTAT online tool[17].

Results

New generation autosomal STR loci *viz.,*D1S1656, D2S1338, D2S441, D3S1358, D5S818, D6S1043,D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, D22S1045, CSF1PO, FGA, TH01, TPOX, vWA, Penta D, Penta E, and SE33 along with sex determination genetic marker Amelogenin, which are available in most recent amplification kits (Table 1);

Table 1. Commercially available most recent and advanced new STR marker based multiplex systems

Sr. No.	Locus	Global Filer® PCR multiplex kit	VeriFiler™ Plus PCR Amplification Kit	PowerPlex® Fusion 6C System	VersaPlex™ 27PY System	Investigator 24plex QS Kit	Ingenomics™ AutoProFiler
1	D3S1358	☑	☑	☑	☑	☑	☑
2	TH01	☑	☑	☑	☑	☑	☑
3	D21S11	☑	☑	☑	☑	☑	☑
4	D18S51	☑	☑	☑	☑	☑	☑
5	Penta E	X	☑	☑	☑	X	☑
6	D12S391	☑	☑	☑	☑	☑	☑
7	D6S1043	X	☑	X	☑	X	☑
8	D2S1338	☑	☑	☑	☑	☑	☑
9	D1S1656	☑	☑	☑	☑	☑	☑
10	D5S818	☑	☑	☑	☑	☑	☑
11	D13S317	☑	☑	☑	☑	☑	☑
12	D7S820	☑	☑	☑	☑	☑	☑
13	D19S433	☑	☑	☑	☑	☑	☑
14	CSF1PO	☑	☑	☑	☑	☑	☑
15	Penta D	X	☑	☑	☑	X	☑
16	D2S441	☑	☑	☑	☑	☑	☑
17	vWA	☑	☑	☑	☑	☑	☑
18	D8S1179	☑	☑	☑	☑	☑	☑
19	TPOX	☑	☑	☑	☑	☑	☑
20	FGA	☑	☑	☑	☑	☑	☑
21	D16S539	☑	☑	☑	☑	☑	☑
22	D22S1045	☑	☑	☑	☑	☑	☑
23	SE33	☑	☑	☑	X	☑	☑
24	D10S1248	☑	☑	☑	☑	☑	☑

and Y STR loci *viz.*, DYS19, DYS385 a/b, DYF387S1 a/b, DYS389 I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS449, DYS456, DYS458, DYS460, DYS481, DYS518, DYS533, DYS549, DYS570, DYS576, DYS627, DYS635 (Y GATA C4), DYS643 and Y GATA H4 were included in this study.The

allele frequency of the genetic data was statistically evaluated for the forensic efficacy parameters. The locus SE33 and Penta E were found as markers with highest degree of forensic efficiency among all the tested loci. The loci D1S1656, Penta E, D18S51, D2S1338, D21S11, D8S1179, D12S391, SE33, and FGA of PowerPlex Fusion 6C™ system kit, were found to be more than 0.8 polymorphic information contents (PIC). The loci D1S1656, Penta E, D18S51, D2S1338, D21S11, D8S1179, D12S391, and FGA of VersaPlex™ 27PY kit, were found to be more than 0.8 polymorphic information contents (PIC). The loci D1S1656, D18S51, D2S1338, D21S11, D8S1179, D12S391, SE33 and FGA of GlobalFiler™ PCR multiplex kit, were found to be more than 0.8 polymorphic information contents (PIC). The loci D1S1656, D18S51, D2S1338, D21S11, D8S1179, D12S391, D13S317, Penta E and FGA of Verifiler Plus™ PCR multiplex kit, were found to be more than 0.8 polymorphic information contents (PIC). All the other forensic efficacy parameters also found to show similar pattern for these markers (Table 2).

Table 2. Forensic efficacy parameters of tested new generation autosomal STR Kits

Forensic efficacy of PowerPlex Fusion 6C system kit (N=374)								Forensic efficacy of VersaPlex™ system (N=147).							
Locus	PD	PIC	PE	PI	Pm	Ho	P-value	locus	PD	PIC	PE	PI	PM	Ho	P-value
D3S1358	0.883	0.700	0.545	2.174	0.117	0.770	0.017	D3S1358	0.881	0.705	0.519	2.042	0.119	0.755	0.170
D1S1656	0.968	0.858	0.664	3.016	0.032	0.832	0.123	D1S1656	0.963	0.866	0.861	7.350	0.037	0.932	0.149
D2S441	0.868	0.649	0.377	1.496	0.132	0.634	0.247	D2S441	0.865	0.665	0.429	1.670	0.135	0.701	0.232
D10S1248	0.907	0.733	0.554	2.226	0.093	0.775	0.164	D10S1248	0.906	0.727	0.519	2.042	0.094	0.755	0.711
D13S317	0.939	0.793	0.569	2.309	0.061	0.783	0.000	D13S317	0.941	0.788	0.495	1.934	0.059	0.741	0.279
Penta E	0.980	0.894	0.674	3.117	0.020	0.840	0.002	Penta E	0.981	0.899	0.777	4.594	0.019	0.891	0.718
D16S539	0.932	0.774	0.588	2.429	0.068	0.791	0.032	D16S539	0.940	0.795	0.604	2.534	0.060	0.803	0.428
D18S51	0.956	0.826	0.669	3.066	0.044	0.837	0.002	D18S51	0.953	0.818	0.630	2.722	0.047	0.816	0.404
D2S1338	0.967	0.859	0.633	2.750	0.033	0.818	0.000	D2S1338	0.967	0.861	0.696	3.341	0.033	0.850	0.371
CSF1PO	0.875	0.665	0.355	1.427	0.125	0.644	0.009	CSF1PO	0.883	0.682	0.440	1.709	0.117	0.707	0.701
Penta D	0.939	0.788	0.503	1.968	0.061	0.746	0.001	Penta D	0.937	0.787	0.484	1.885	0.063	0.735	0.198
THO1	0.910	0.729	0.429	1.670	0.090	0.578	0.001	TH01	0.902	0.726	0.462	1.793	0.098	0.721	0.066
vWA	0.940	0.790	0.545	2.174	0.060	0.770	0.032	vWA	0.930	0.770	0.554	2.227	0.070	0.776	0.819
D21S11	0.959	0.829	0.613	2.597	0.041	0.778	0.691	D21S11	0.947	0.823	0.736	3.868	0.053	0.871	0.176
D7S820	0.924	0.762	0.564	2.280	0.076	0.781	0.349	D7S820	0.927	0.763	0.530	2.100	0.073	0.762	0.376
D5S818	0.868	0.673	0.485	1.889	0.132	0.735	0.042	D5S818	0.866	0.668	0.440	1.709	0.134	0.707	0.407
TPOX	0.843	0.619	0.393	1.545	0.157	0.676	0.403	TPOX	0.864	0.652	0.306	1.289	0.136	0.612	0.062

D8S1179	0.956	0.829	0.613	2.597	0.044	0.807	0.010
D12S391	0.961	0.838	0.623	2.671	0.039	0.805	0.005
D19S433	0.939	0.781	0.554	2.226	0.061	0.722	0.000
SE33	0.990	0.940	0.770	4.452	0.010	0.877	0.000
D22S1045	0.881	0.673	0.313	1.308	0.119	0.618	0.000
FGA	0.974	0.873	0.649	2.877	0.026	0.821	0.108

D8S1179	0.948	0.813	0.630	2.722	0.052	0.816	0.764
D12S391	0.957	0.841	0.669	3.063	0.043	0.837	0.311
D19S433	0.936	0.780	0.554	2.227	0.064	0.776	0.636
D6S1043	0.934	0.775	0.591	2.450	0.066	0.796	0.686
D22S1045	0.877	0.675	0.440	1.709	0.123	0.707	0.415
FGA	0.968	0.864	0.709	3.500	0.032	0.857	0.509

Forensic efficacy of GlobalFiler PCR kit (N=508)

Locus	PD	PIC	PE	PI	Pm	Ho	P-value
D3S1358	0.887	0.696	0.470	1.827	0.113	0.726	0.406
vWA	0.939	0.794	0.653	2.920	0.061	0.829	0.490
D16S539	0.934	0.772	0.583	2.396	0.066	0.791	0.282
CSF1PO	0.867	0.662	0.439	1.705	0.133	0.707	0.228
TPOX	0.865	0.652	0.397	1.558	0.135	0.679	0.010
D8S1179	0.950	0.818	0.707	3.479	0.050	0.856	0.255
D21S11	0.957	0.833	0.771	4.456	0.043	0.888	0.471
D18S51	0.956	0.826	0.731	3.791	0.044	0.868	0.750
D2S441	0.863	0.686	0.583	2.396	0.137	0.791	0.013
D19S433	0.929	0.778	0.676	3.136	0.071	0.841	0.400
TH01	0.905	0.730	0.587	2.419	0.095	0.793	0.168
FGA	0.960	0.846	0.803	5.184	0.040	0.904	0.012
D22S1045	0.879	0.684	0.500	1.954	0.121	0.744	0.730
D5S818	0.878	0.692	0.527	2.082	0.122	0.760	0.272
D13S317	0.935	0.780	0.587	2.419	0.065	0.793	0.061
D7S820	0.930	0.768	0.594	2.466	0.070	0.797	0.788
SE33	0.991	0.939	0.847	6.684	0.009	0.925	0.014
D10S1248	0.904	0.740	0.665	3.024	0.096	0.835	0.087
D1S1656	0.967	0.853	0.680	3.175	0.033	0.843	0.212
D12S391	0.957	0.830	0.696	3.342	0.043	0.850	0.493
D2S1338	0.964	0.849	0.711	3.528	0.036	0.858	0.085

Forensic efficacy parameters of of VeriFiler Plus (N=200)

Locus	PD	PIC	PE	PI	Pm	Ho	P-value
D3S1358	0.886	0.699	0.510	2.000	0.114	0.750	0.372
vWA	0.934	0.775	0.599	2.500	0.066	0.800	0.897
D16S539	0.920	0.756	0.581	2.381	0.080	0.790	0.706
CSF1PO	0.878	0.683	0.484	1.887	0.122	0.735	0.101
D6S1043	0.937	0.788	0.646	2.857	0.063	0.825	0.081
D8S1179	0.956	0.838	0.715	3.571	0.044	0.860	0.185
D21S11	0.955	0.833	0.675	3.125	0.045	0.840	0.014
D18S51	0.950	0.819	0.675	3.125	0.050	0.840	0.362
D5S818	0.873	0.679	0.493	1.923	0.127	0.740	0.707
D2S441	0.899	0.708	0.527	2.083	0.101	0.760	0.710
D19S433	0.931	0.775	0.590	2.439	0.069	0.795	0.751
FGA	0.958	0.840	0.745	4.000	0.042	0.875	0.494
D10S1248	0.913	0.744	0.590	2.439	0.087	0.795	0.732
D22S1045	0.885	0.691	0.413	1.613	0.115	0.690	0.206
D1S1656	0.973	0.880	0.816	5.556	0.027	0.910	0.301
D13S317	0.946	0.805	0.581	2.381	0.054	0.790	0.284
D7S820	0.929	0.769	0.562	2.273	0.071	0.780	0.318
Penta E	0.982	0.908	0.795	5.000	0.018	0.900	0.185
Penta D	0.935	0.785	0.627	2.703	0.065	0.815	0.870
TH01	0.921	0.756	0.562	2.273	0.079	0.780	0.970
D12S391	0.959	0.840	0.627	2.703	0.041	0.815	0.255
D2S1338	0.968	0.867	0.775	4.545	0.032	0.890	0.536
TPOX							
0.849							
0.656							
0.468							
1.818							
0.151							
0.725							
0.026							

*Abbreviations: PD – power of discrimination; PIC – polymorphic information content; PE – power of exclusion; PI – paternity index; PM – matching probability; Ho - observed heterozygosity; P-value- deviation from Hardy–Weinberg equilibrium

The results of present study reveal following genetic marker namely SE33, Penta E, FGA,D1S1656, D18S51, D2S1338, D21S11, D8S1179, D12S391 and D13S317 must be included in the multiplex system used for forensic DNA application for the Indian population. Studied Y chromosomal STR markers showed significant genetic diversity among the analyzed male populations. Therefore, 23 or more Y-STR genetic markers are sufficient for the forensic application.

Discussion

Since the advent of Indian forensic set-up in the late 1980s application of DNA based approaches in forensic analysis have taken center stage. Marker analysis and choice of markers for analysis of DNA samples on Indian population relied heavily on the American or European standards, this has limited to a certain degree the analysis and population forensic study aspects. Scattered scientific reports, coming out of research from Indian scientific community, have slowly started to shed light on certain Autosomal STR and Y chromosomal STRs and their significance in sample analysis. The present study prompted to look at these markers in greater depth to further draw their statistical relevance and mandatory selection in DNA profiling. The study elaborated on application of almost 25 Autosomal STR markers and ~23 or more Y chromosomal STR markers. Analysis has vividly indicated the short falls in choosing exclusively either CODIS loci or ENFSS loci. A broader look at the statistically driven data on the STR markers relevant to Indian Population indicated that SE33, Penta E, FGA, D1S1656, D18S51, D2S1338, D21S11, D8S1179, D12S391 and D13S317 play a significant role is case work analysis with a very high degree of forensic efficacy. These outcomes categorically point to the need of scientific community to give careful consideration in selecting multiplex systems that include these markers (ISDMS) for their studies related to forensic DNA testing, DNA Data basing and Population Genetics. The NDIS (National DNA Index System) is developed solely for the American database development and so does the CODIS (Combined DNA Index System). Also, the European DNA Profiling Group (EDNAP), and European Standard Set (ESS) are developed with importance in the populations related to the continent of Europe. Lack of concrete analysis on Indian marker set drew this study to initiate the development of Indian Standard DNA marker System (ISDMS), a much-needed system to help Indian forensic community.

Conclusion

While selecting the Multiplex systems for Forensic DNA analysis it becomes imperative that laboratories chose kits that include the above listed markers viz., SE33, Penta E, FGA, D1S1656, D18S51, D2S1338, D21S11, D8S1179, D12S391 and D13S317, that bear relevance in Indian population. With a diverse population as Indian, choosing marker systems that are unrelated may lead analysts to miss-out on critical information which otherwise would serve beneficially in solving complex forensic cases. In this study we also for the first time introduced the acronym "ISDMS (Indian Standard DNA marker System)" which can be further quoted while discussing about these and any additional markers that will be added to the list of markers specific to Indian Population.

Acknowledgements

Authors thankfully acknowledge Promega (India) and Thermo Fisher Scientific for providing the multiplex STR kit for the study.

Conflict of interest statement

The authors declare that they have no conflict of interest.

References

[1] J.M. Butler, Forensic DNA typing: biology, technology, and genetics of STR markers, Elsevier, 2005.

[2] P. Gill, A.J. Jeffreys, D.J. Werrett, Forensic application of DNA 'fingerprints,' Nature 318 (1985) 577.

[3] A.J. Jeffreys, V. Wilson, S.L. Thein, Individual-specific 'fingerprints' of human DNA, Nature 316 (1985) 76–79.

[4] P. Shrivastava, T. Jain, R.K. Kumawat, Direct PCR amplification from saliva sample using non-direct multiplex STR kits for forensic DNA typing, Sci. Rep. 11 (2021) 1–19.

[5] R. Giroti, I. Talwar, Diversity and differentiation in Khatris, Banias and Jat Sikhs of Punjab: a study with forensic microsatellites, Ind J Phys Anthr. Hum Genet 32 (2013) 309–328.

[6] R.K. Kumawat, P. Shrivastava, D. Shrivastava, G.K. Mathur, S. Dixit, Genomic blueprint of population of Rajasthan based on autosomal STR markers, Ann. Hum. Biol. (2020) 1–6.

[7] P. Shrivastava, T. Jain, V. Ben Trivedi, Genetic polymorphism study at 15 autosomal locus in central Indian population, Springerplus 4 (2015) 566.

[8] M.J. Ludeman, C. Zhong, J.J. Mulero, R.E. Lagacé, L.K. Hennessy, M.L. Short, D.Y. Wang, Developmental validation of GlobalFiler™ PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples, Int. J. Legal Med. 132 (2018) 1555–1573. https://doi.org/10.1007/s00414-018-1817-5.

[9] M. Bamshad, T. Kivisild, W.S. Watkins, M.E. Dixon, C.E. Ricker, B.B. Rao, J.M. Naidu, B.V.R. Prasad, P.G. Reddy, A. Rasanayagam, Genetic evidence on the origins of Indian caste populations, Genome Res. 11 (2001) 994–1004.

[10] R.K. Kumawat, P. Shrivastava, D. Shrivastava, G.K. Mathur, H.R. Dash, G. Chaubey, Peopling of Rajasthan, India: Evaluating the gene flow from east and west, Gene Reports 22 (2021) 100990.

[11] P.P. Rickham, Human experimentation. Code of ethics of the world medical association. Declaration of Helsinki., Br. Med. J. 2 (1964) 177.

[12] S. Dixit, P. Shrivastava, R.K. Kumawat, K. Kaitholia, H.R. Dash, H. Sharma, G. Choubey, Forensic genetic analysis of population of Madhya Pradesh with PowerPlex Fusion 6C™ Multiplex System, Int. J. Legal Med. (2019) 1–3.

[13] P. Shrivastava, S. Dixit, R.K. Kumawat, A. Srivastava, Efficiency analysis of VersaPlex™ 27PY system in central Indian population: first report from Indian population, Leg. Med. 54 (2022) 101983.

[14] S. Sahoo, R. Samal, S. Behera, A.K. Swain, S. Biswas, P. Shrivastava, R.K. Kumawat, G. Chaubey, H.R. Dash, Genomic portrait of Odisha, India drawn by using 21 autosomal STR markers, Int. J. Legal Med. (2020) 1–3.

[15] P. Shrivastava, K. Kaitholia, R.K. Kumawat, S. Dixit, H.R. Dash, A. Srivastava, Y. Baroniya, A. Mishra, P. Kushwaha, M. Rana, Forensic effectiveness and genetic distribution of 23 autosomal STRs included in Verifiler Plus TM multiplex in a population sample from Madhya Pradesh, India, Int. J. Legal Med. (2019) 1–2.

[16] R.O.D. Peakall, P.E. Smouse, GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research, Mol. Ecol. Resour. 6 (2006) 288–295.

[17] P.G. Ristow, M.E. D'Amato, Forensic statistics analysis toolbox (FORSTAT): A streamlined workflow for forensic statistics, Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) e52–e54.

Table Legends:

Table 1. Commercially available most recent and advanced new STR marker based multiplex systems

Table 2. Forensic efficacy parameters of tested new generation autosomal STR Kits

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 603-610
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Genetic relationships of semi-feral subpopulations of Ecuadorian highland creole horse or "Paramero" (Caballo Paramero Ecuatoriano)

Jorge Navarrete-Mera[1*], Viviana Pazmiño[2], Marco Coral[1,3], Camila Lara[2], Diego Alarcón[1], Andrés Ambuludí[2,4], Katherin Barrionuevo-Pérez[2,5], Amparo Martínez-Martínez[6], José Luis Vega-Pla[7] and Germán Burgos[8,9*]

[1]*Escuela de Medicina Veterinaria, Facultad de Ciencias de la Salud. Universidad de las Américas (UDLA), Quito-Ecuador.*

[2]*Carrera de Biotecnología, Universidad de Las Américas (UDLA), Quito, Ecuador.*

[3]*Grupo de Bio-quimioinformática. Universidad de las Américas (UDLA), Quito-Ecuador.*

[4]*Laboratorio de Biología Molecular, Agencia de Regulación y Control Fito y Zoosanitario AGROCALIDAD, Quito, Ecuador*

[5]*Laboratorios de Investigación y Vinculación. Universidad de las Américas (UDLA), Quito-Ecuador*

[6]*Departamento de Genética. Universidad de Córdoba, Campus Universitario de Rabanales, Córdoba-España*

[7]*Laboratorio de Investigación Aplicada, Cría Caballar de las Fuerzas Armadas. Córdoba-España.*

[8]*One Health Research Group, Facultad de Medicina, Universidad de Las Américas (UDLA), Quito, Ecuador.*

[9]*Grupo de Medicina Xenómica, Universidad de Santiago de Compostela, Santiago de Compostela, España.*

** Corresponding author at Germán Burgos, Universidad de Las Américas, campus UDLAPark, Via Antigua a Nayon, Quito-Ecuador. Email hgermanburgosf@yahoo.com. Tel +593 995256888 fax +593 23970000*

Abstract

The Highland Ecuadorian Creole horse (HECH), also known as Paramero, originated from the introduction of European equines by the Spanish in the 16th century. These horses have since adapted to the high-altitude environments of the Ecuadorian Andes, particularly in areas known as "paramos" at elevations above 3300 meters. In this study, we aimed to genotype and analyze the genetic diversity of feral HECH populations from four provinces in Ecuador using 29 STR molecular markers. DNA was extracted from 182 horses across the provinces of Carchi, Pichincha, Cotopaxi, and Chimborazo, and genetic diversity parameters, F-statistics, and population

structure were assessed. The results revealed high levels of genetic diversity across all populations, with little genetic differentiation (FST=0.01007) between subpopulations. Bayesian clustering analysis indicated that HECH grouped with Creole and Iberian breeds, and at K=11, they formed a distinct cluster. The genetic affinity of HECH with other Creole and Iberian populations suggests a common origin from southern Spain, likely introduced to Ecuador through Central America. These findings highlight the genetic richness of HECH and provide a foundation for future conservation and breeding programs.

Keywords
Horse, Ecuadorian, Paramero-Creole, STRs, Genetic diversity, Feral-subpopulation.

Introduction

Highland Ecuadorian Creole horses (HECH), also known as Paramero, originated from the introduction of European equines by the Spanish in the 16th century [1]. In the Ecuadorian Andes, HECH adapted to conditions from high altitude areas known as "paramos" located at elevations of approximately 3300 meters above sea level [2]. HECH is deeply rooted in Andean communities´s lives and traditions, consider HECH as part of their culture. Semi-wild herds of HECH roam the paramo and are regularly managed by horse riders of local communities [3]. This study aimed to genotype and compare feral populations of HECH horses from four Andean provinces of Ecuador using microsatellite markers to assess their diversity and relationships and give valuable information for their conservation.

Material studied, methods, techniques

Material & Methods

DNA was extracted from the blood of 182 feral horses from herds in the paramos of the provinces of Carchi (n=40), Pichincha (n=61), Cotopaxi (n=50), and Chimborazo (n=33). STR genotypes were obtained using 29 molecular markers, following the protocols described in [4,5]. PCR products were run on an ABI 3500 Genetic analyzer and genotypes were edited in GeneMapperTM v6, including two independent lectures.

Genetic analysis

Allelic frequencies and other diversity parameters were computed using GenAlEx 6.5 [6], forensic parameters using Powerstats software, [7] and F-statistics

were estimated using Genetix 4.05.2 [8]. HW disequilibrium was assessed by an exact test in GENEPOP 7 [9]. Nei's standard genetic distance was calculated with reference population data from the BIOHORSE consortium using Populations 1.2.32 software [10] and an unrooted Neighbor-Joining tree was plotted with TreeViewer 2.2.0 [11]. Analysis for population structuration was made using Structure v2.3.4 software [12].

Results

All four populations exhibited high genetic diversity and heterozygosity values. F-Statistics indicated small variability between HECH populations (FIT=0.05927; CI:0.02828-0.11244), and the horse subpopulations were relatively homogeneous (FIS=0.04970; CI:0,01835-0,10354), with no significant genetic differentiation observed (FST=0.01007; CI:0,00714-0,01310). This result suggests that the HECH populations studied are genetically similar, which is consistent with the historical spread of horses across Eurasia before their introduction to the Americas during the European conquest [13–15].

Among the markers analyzed, T333, CA425, T312, ASB17, HTG10, T297, T394, ASB23, T343, and T325 were identified as the most efficient. In contrast, HMS6 and LEX33 showed Hardy-Weinberg disequilibrium and were the least efficient markers in the studied populations (Table a).

Table a. *Diversity and forensic parameters of Paramero Horses as a single population.*

Parameter/Locus	VHL20	HTG4	AHT4	HMS7	ASB2	ASB17	HMS6	AHT5	ASB23	HTG10	HMS3	LEX33	T287	T294	T297	T301	T312	T321	T325	T333	T337	T341
Na	10	8	9	8	10	17	11	9	11	10	8	9	11	7	11	8	11	10	11	10	7	10
Ne	5,523	3,157	5,959	3,636	5,100	7,199	3,530	5,157	6,166	6,247	5,777	6,177	3,865	3,150	6,728	3,862	6,268	3,739	7,067	7,007	3,988	4,297
oHe	0,777	0,707	0,772	0,685	0,745	0,842	0,663	0,804	0,832	0,842	0,766	0,299	0,738	0,661	0,835	0,732	0,857	0,720	0,809	0,857	0,665	0,747
uHe	0,821	0,685	0,834	0,727	0,806	0,863	0,719	0,808	0,840	0,842	0,829	0,840	0,743	0,684	0,854	0,743	0,843	0,735	0,861	0,860	0,751	0,769
PIC	0,795	0,652	0,813	0,698	0,777	0,848	0,683	0,780	0,817	0,822	0,805	0,819	0,704	0,629	0,834	0,708	0,825	0,706	0,843	0,841	0,716	0,737
CP	0,056	0,134	0,048	0,107	0,070	0,035	0,121	0,063	0,051	0,048	0,056	0,089	0,108	0,160	0,042	0,109	0,044	0,102	0,040	0,042	0,099	0,092
PD	0,944	0,866	0,952	0,893	0,930	0,965	0,879	0,937	0,949	0,953	0,944	0,911	0,892	0,840	0,958	0,891	0,956	0,898	0,960	0,958	0,901	0,908
PE	0,557	0,438	0,548	0,405	0,501	0,680	0,373	0,607	0,659	0,680	0,538	0,063	0,489	0,371	0,666	0,480	0,709	0,460	0,615	0,709	0,376	0,505
TPI	2,244	1,704	2,191	1,586	1,957	3,172	1,484	2,556	2,968	3,172	2,140	0,713	1,906	1,476	3,033	1,867	3,500	1,784	2,614	3,500	1,492	1,978
MAF	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007	0,007
HWE	0,560	0,684	0,163	0,683	0,037	0,604	0,001[a]	0,812	0,465	0,527	0,013	0,000[a]	0,755	0,062	0,482	0,013	0,204	0,660	0,064	0,226	0,008	0,054

Na= Number of Alelles; Ne= number of effective Alelles; oHe= Observed Heterozygosity; uHe= Unbiased Expected Heterozygosity; PIC= Polymorphism Informative Content; CP= Coincidence Probability; PD= Power of discrimination; PE= Power of Exclusion; TPI= Typical Paternity Index; MAF= Minimal Allele Frequency; HWE= Hardy-Weinberg Equilibrium by exact text; [a] = Significative after Bonferroni correction p-value=0.001724.

The Neighbor-Joining dendrogram, based on Nei's standard genetic distance, revealed a close genetic relationship between HECH and other Creole breeds, such as the Colombian and Peruvian "Paso Fino," Salvadoran Creole, Pantaneiro from Brazil (with Andalusian ancestry), and Panamanian Creole. The short branch length of HECH in the dendrogram suggests limited evidence of directed selection compared to other breeds such as Thoroughbred, Retuertas, and Argentinian Creole (Figure 1).

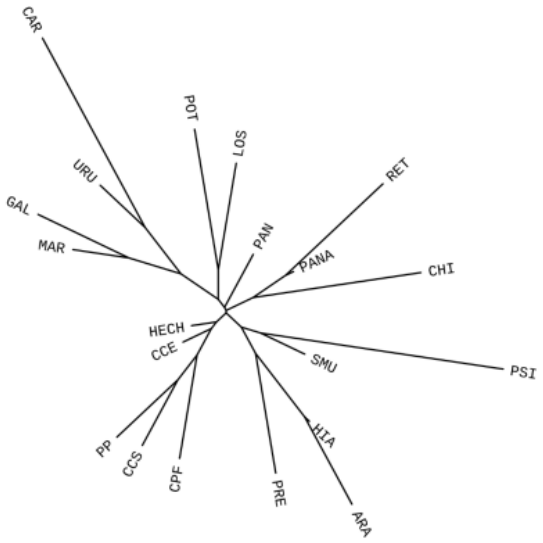


Figure 1. Neighbor-joining dendrogram based on Nei's Standard genetic distances of 20 horse populations.

Bayesian genetic structure analysis showed that HECH clusters with other Creole and Iberian breeds (K = 2, K = 3). At K = 11, HECH populations formed a distinct, uniform group, even when compared with the other Creole Ecuadorian populations from the BioHorse Consortium (Figure 2).

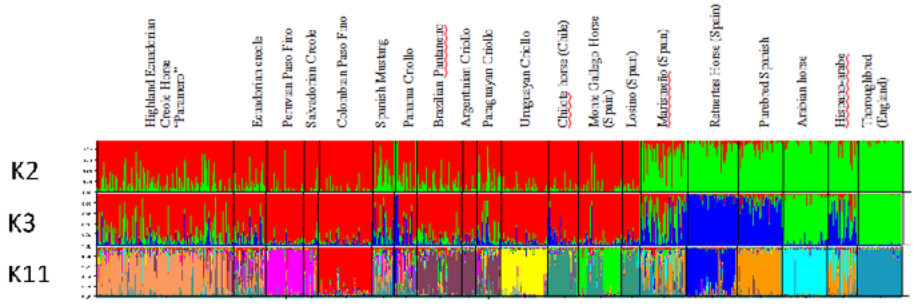


Figure 2. *Bayesian clustering for population structuration for 20 breeds of populations.*

Discussion

The high genetic diversity and heterozygosity observed across all four HECH populations are consistent with findings in other Creole horse breeds, which are often not subject to intense selective breeding. These high diversity values are crucial for the long-term viability of the population, providing a broad genetic pool that may help these horses adapt to environmental changes, such as those associated with high-altitude paramo conditions.

The low levels of genetic differentiation between the populations (FST = 0.01007) suggest a lack of significant barriers to gene flow among the HECH populations in the four provinces. This finding aligns with the historical context of horse migration and population admixture following their introduction to the Americas. The relatively homogeneous structure seen in the HECH population supports the idea that these horses share a common ancestry and can be treated as a single genetic unit in conservation and breeding programs.

The identification of efficient and inefficient markers has important implications for future genetic studies in these populations. The markers that showed Hardy-Weinberg disequilibrium, HMS6, and LEX33, should be revisited or potentially excluded in future analyses to avoid misleading interpretations of genetic variability.

The close genetic relationship between HECH and other Creole breeds, such as those from Colombia, Peru, El Salvador, Brazil, and Panama, highlights the shared Iberian ancestry of these horses. The short branch length of HECH in the dendrogram indicates a lack of intense artificial selection, distinguishing HECH from more heavily selected breeds like Thoroughbred, Retuertas, and Argentinian Creole. This lack of directed selection could contribute to the preservation of genetic diversity in the HECH population, further supporting its conservation.

The Bayesian clustering analysis reinforces these conclusions by showing a strong genetic grouping of HECH with Creole and Iberian breeds at lower K values (K = 2, K = 3). The formation of a distinct group at K = 11 suggests that HECH has retained a unique genetic signature within the broader Creole horse population, emphasizing the need to manage HECH as a distinct genetic resource. Additionally, the Chimborazo population's slight genetic isolation, as shown in the clustering at higher K values, could indicate the beginning of localized genetic differentiation, warranting further investigation.

Conclusion

HECH or Paramero populations exhibit high genetic diversity. In addition, in these populations, the set of markers is highly polymorphic and powerful for parentage testing.

F-statistics results indicate that the four populations are not genetically different, the Bayesian clustering in Structure can be confirmed this, and therefore, can be considered as one population.

Structure and NJ tree results demonstrate a genetic affinity between HECH and certain Creole and Iberian breeds, supporting their arrival to Ecuador through Central America and originating mainly from southern Spain.

Acknowledgments

To the "Quito-Ecuestre" association for their decisive initiative and financial support for the characterization of the Paramero Horses and for their valuable help in organizing and executing the travels for sampling in the Ecuadorian highlands. UDLA authors (JN-M, VP, MC, CL, DA, AA, KB-P, and GB) were funded for the laboratory analysis by MED.GBF.20.01 grant from "DGIV - Universidad de Las Américas, Quito-Ecuador"

Conflict of interest statement

None

References

[1] P. Mitchell, I. Rode, Equestrian Adaptations of Indigenous Peoples in Southern Hemisphere Arid Zones, Int J Hist Archaeol 21(2017) 321–345. https://doi.org/10.1007/s10761-016-0380-3.

[2] S.L. Hendrickson, A genome-wide study of genetic adaptation to high altitude in feral Andean Horses of the páramo, BMC Evol Biol 13(2013) 273. https://doi.org/10.1186/1471-2148-13-273.

[3] C.O. Larrea-Izurieta, F.J. Oñate-Mancero, M.P. Paredes-Orozco, Estudio Zoométrico de caballos criollos parameros ecuatorianos en la provincia de Chimborazo, Ecuador, Revista Científica de La Facultad de Ciencias Veterinarias de La Universidad de Zulia XXVIII (2018) 257–264.

[4] T. Tozaki, H. Kakoi, S. Mashima, et al., Population Study and Validation of Paternity Testing for Thoroughbred Horses by 15 Microsatellite Loci, 2001.

[5] FAO, Molecular genetic characterization of animal genetic resources, (2011). https://www.fao.org/3/i2413e/i2413e.pdf (accessed September 9, 2024).

[6] R. Peakall, P.E. Smouse, GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update, Bioinformatics 28(2012) 2537–2539. https://doi.org/10.1093/bioinformatics/bts460.

[7] Promega Co., PowerStats: a computer program for the analysis of population statistics, Free Program Distributed by the Authors over the Internet, (1999). http://www.promega.com/geneticidtools. (accessed July 30, 2024).

[8] K. Belkhir, GENETIX, logiciel sous WindowsTM pour la génétique des populations, http://www.Univmontp2. Fr/~ Genetix (1999).

[9] F. Rousset, Genepop'007: a complete re-implementation of the genepop software for Windows and Linux, Mol Ecol Resour 8(2008) 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x.

[10] O. Langella, Populations 1.2. 32. Population Genetic Software. Gif-sur-Yvette, France: Laboratoire Evolution, Génomes et Spéciation, (2015).

[11] R.D.M. Page, Visualizing phylogenetic trees using TreeView, Curr Protoc Bioinformatics (2003) 2–6.

[12] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of Population Structure Using Multilocus Genotype Data, Genetics 155(2000) 945–959. https://doi.org/10.1093/genetics/155.2.945.

[13] S. Lippold, N.J. Matzke, M. Reissmann, et al., Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication, BMC Evol Biol 11(2011). https://doi.org/10.1186/1471-2148-11-328.

[14] C. Vilà, J.A. Leonard, A. Götherström, et al., Widespread Origins of Domestic Horse Lineages, Academic Press, 2000. www.sciencemag.org.

[15] A. Fages, K. Hanghøj, N. Khan, et al., Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series, Cell 177(2019) 1419-1435. e31. https://doi.org/10.1016/j.cell.2019.03.049.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 611-618
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Age Prediction from Semen Samples via DNA Methylation Analysis after Quality Control of Bisulfite Treated DNA

Lucie Kotková[1], Jana Vrbková[1], Rastislav Slavkovský[1], Eliška Remešová[1], Helena Jurtíková[1], Barbora Líznerová[1], Jiří Drábek[1*]

[1]*Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University and University Hospital Olomouc, Czech Republic*

** Corresponding author at: Institute of Molecular and Translational Medicine, Hněvotínská 5, Olomouc, 77900, Czech Republic, jiri.drabek@upol.cz, +420585632070*

Abstract

Age prediction remains one of the foremost pursuit tools in forensic DNA phenotyping. The current standard methods detect cytosine methylation via pyrosequencing or massively parallel sequencing after bisulfite conversion, whose harsh and destructive nature poses limitations in low quality and/or quantity DNA analyses. The absence of a universally accepted method for evaluating the quality of bisulfite conversion presents a challenge, as incomplete conversion can significantly distort downstream results.

We opted to employ the BisQuE method, a multiplex quantitative real-time PCR technique for bisulfite conversion quality control developed by Hong and Shin, and to develop an age prediction model for semen samples in the Czech population based on massively parallel sequencing.

Our age prediction model for semen samples utilizes six CpGs in six genes (*SH2B2, SYT7, NOX4, TTC7B, TBX4,* and *GALR2*) and reaches mean absolute error (MAE) 2.62 years. This model was constructed using a dataset comprising 46 samples (35 volunteers from a fertility clinic and 11 healthy volunteers, 19-52 years old). Given the comparatively low number of samples, we tested the possibility to include sperm samples exposed to cryoprotectives. We also focus on possible differences of methylation profile between ejaculate and its fractions.

Here, we present our results of the BisQuE workflow, including modification of the original protocol by incorporation of salmon sperm DNA.

We also present the performance characteristics of age prediction model MethAge for sperm samples, along with the results of testing the effect of cryoprotective medium and differences between the methylation profiles from sperm and non-sperm fractions of ejaculate in the pilot sample set (n = 5).

Keywords
Epigenetics, age prediction, semen, differential extraction, bisulfite conversion, and BisQuE.

Introduction

The MethAge is a massively parallel sequencing method for age prediction from semen samples, suitable for sexual assault cases, developed for the Czech population. MethAge was primarily developed for use with full ejaculate samples. Still, as ejaculate is a highly heterogenous material (both between individuals and in time), we also tested its performance with separate sperm and non-sperm fractions after differential extraction (Differex IQ system, Promega). Differential extraction is often necessary (to eliminate female DNA in vaginal swabs) but can be redundant (e.g., semen stains on fabric) and the performance of the model can be different for both materials.

Semen is also considered very private material, limiting sample availability. Using samples frozen in a cryoprotective medium could significantly expand the sample pool, but some studies suggest that cryoprotectives disrupt methylation signals, though results are sometimes contradictory based on the used cryoprotectant agents, analyzed loci, and tested species [1].

MethAge method includes bisulfite conversion (EZ Methylation Gold, Zymo), qPCR, library preparation (Nextera XT, Illumina), massively parallel sequencing (MiSeq/NovaSeq Illumina), and bioinformatic analysis (Bismarck and SeqMonk). Bisulfite conversion step seems to be a bottleneck of the method that potentially distorts results by induced DNA degradation and loss. Without a universally accepted method for evaluating conversion quality (quantifying conversion efficiency, DNA degradation, and DNA loss), we decided to implement the BisQuE method, a multiplex quantitative real-time PCR technique developed by Hong and Shin [2].

This study aims to assess the MethAge performance for semen samples and compare the age prediction results from whole semen, sperm and non-sperm DNA fractions after differential extraction and DNA stored in cryoprotective medium PreservCyt. Concurrently, we also aim to implement the BisQuE method, including verification of its ability to identify incomplete bisulfite conversion and to test the conversion efficiency of our MethAge workflow.

Material studied, methods, techniques

Samples

The MethAge model was developed using 46 samples with informed consent (35 samples from the Fertility Clinic of University Hospital Olomouc and 11 samples from healthy volunteers from general population).

Upon arrival, all samples were stored in -20 °C without any additives. For 11 samples with sufficient volume, 150-600 ml of sample was aliquoted and stored for two years in -20 °C with addition of 500 ml of cryoprotective medium PreservCyt.

DNA extraction

DNA extraction from full semen was performed using Genomic Mini kit (A&A Biosystems) from 100 ml of semen according to manufacturer instruction. Prior to extraction, samples were allowed to thaw in the room temperature and mixed by pipetting.

Differential extraction was performed using Differex system (Promega) from 100 ml of semen according to manufacturer instruction for differential extraction using the Differex magnet. DNA extraction from individual fractions was subsequently performed using DNA IQ system (Promega). For the DNA extraction from non-sperm cell fraction, 300 ml was used.

MethAge

The MethAge prediction workflow can be seen on the figure 1. Specifications of individual steps are beyond the scope of this short communication. They are currently being prepared for independent publication and are available from the authors upon request.

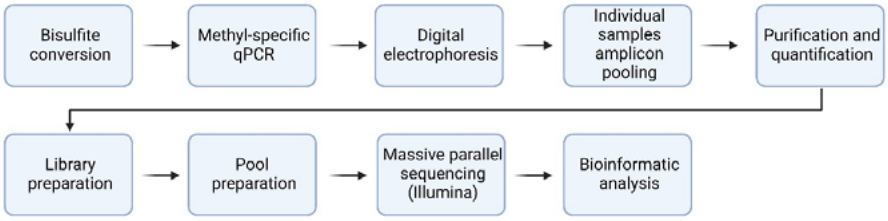


Figure 1. MethAge workflow.

During the development of the MethAge model, we tested 41 CpG loci in 11 genes associated with chronological ageing in semen samples according to literature. Using LASSO selection, stepwise selection, and Bayes modelling, we selected 6 CpG loci in 6 genes (SH2B2, SYT7, NOX4, TTC7B, TBX4, and GALR2) for the final model.

BisQuE

The BisQuE method incorporates 3 sets of primers for amplicons of various lenghts and 4 fluorescent probes. Short amplicon (*CCDC29* gene, 104 bp) incorporates the C base. Its conversion status is detected using short-C and short-T probes (FAM and VIC/HEX, respectively). Long amplicon (*FHJ39739* gene, 238 bp) that does not contain any C base is amplified regardless of conversion status and fluorescent signal from long C-free probe (NED/ROX) is used for fragmentation assessment. Artificially prepared amplicon (147 bp) and its detection probe (Cy5) are used as internal control of qPCR reaction.

For calibration of results, the dilution series of genomic DNA standard and dilution series of synthetic C- and T- indicator containing degenerate base Y needs to be included in each BisQuE run. For each sample, gDNA and bisulfite-converted DNA (BK-DNA) are recommended to run in duplicate.

Results

During the implementation phase of the BisQuE method, we tested three different polymerase-buffer systems and achieved the highest probe specificity using Platinum II Hot-Start PCR Master Mix (Invitrogen).

We experienced repeated failures of dilution series of DNA standards and indicators, that made the whole BisQuE run unanalyzable. The same dilution series samples were simultaneously tested using another qPCR system (GAPHD gene and SYBR Green I probe, routinely used in our lab) with satisfactory results, so the failure was attributed to the BisQuE workflow. The addition of salmon sperm DNA and letting prepared dilution series sit in a fridge for 24 hours before the run lowered but not eliminated the number of failed reactions (see table 1). On the other hand, the addition of salmon sperm DNA limited the unspecific reactions in the non-template control.

Table 1. R^2 values of short C, long and T->C transforming qPCR curve constructed using dilution series of DNA standards and C and T indicators. Values come from independent experiments with addition of salmon sperm DNA. T_0 values were obtained from run that was started immediately after preparation of dilution series, for T_{24} values, dilution series were allowed to sit in the fridge for 24 hours. Optimal value should be > 0.99.

	Experiment 1		Experiment 2		Experiment 3		Experiment 4		Experiment 5	
	T0	T24	T0	T24	T0	T24	T0	T24	T0	T24
Short C	0.993	0.988	0.998	0.996	N/A	N/A	N/A	N/A	0.951	0.867
Long	0.997	0.998	0.992	0.990	0.940	0.977	0.994	0.982	0.937	0.943
T->C	1.000	0.982	0.992	0.994	0.870	0.813	0.971	N/A	0.988	0.799

Two samples with artificially inefficient conversion (mixture of gDNA and converted DNA) were tested to assess the capacity of BisQuE to detect incomplete bisulfite conversion. All our tested samples (n = 20) show 100% conversion efficiency and artificially inefficient conversion was identified as such *(see the figure 2)*

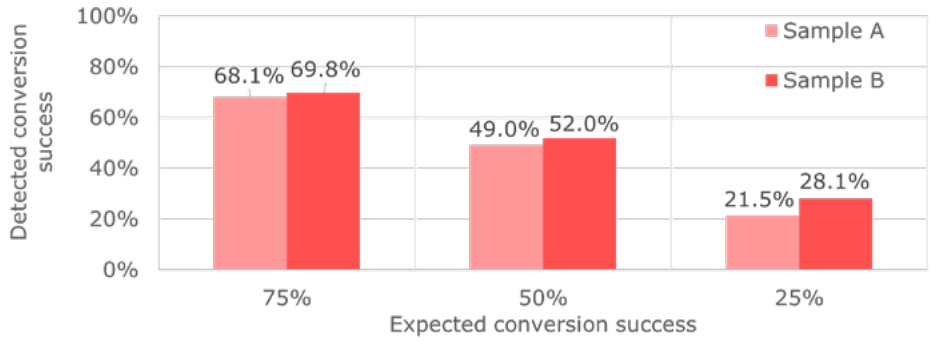


Figure 2. Results of BisQuE conversion efficiency testing in samples with incomplete conversion.

The fragmentation analysis results showed no correlation with digital electrophoresis (D1000 Screen Tape system, Agilent and 2100 Bioanalyzer, Agilent). The original formula for DNA recovery value resulted in values exceeding 100% [2]. *Alternative formula suggested by authors determined the recovery values for two illustrative samples as 16.3% and 9.3%, compared with recovery assessed by Qubit as 73.0% and 83.6%, respectively.*

Age prediction from the full semen samples (n = 46) using the MethAge model resulted in the 82.6% of samples in the range of ± 4 years and in the 89.1% of samples in the range of ± 5 years with the MAE of 2.62 years.

Preliminary testing of 5 samples after differential extraction show that the errors of sperm fraction are higher than the errors of non-sperm fraction (see table 2). We can also see that cryoprotective medium did not affect the prediction accuracy.

Table 2. Mean errors of age prediction from the whole semen, from fractions after differential extraction, and from sperm stored in cryoprotective medium.

	S1	S2	S3	S4	S5
Whole semen	2.46	6.34	6.14	3.02	7.24
Sperm fraction	1.20	12.19	7.21	5.23	6.00
Non-sperm fraction	0.04	0.20	2.88	4.70	6.72
PreservCyt stored	1.80	6.05	6.56	5.33	6.52

Discussion

We developed an epigenetic-based age prediction model called MethAge applicable to the semen samples with mean absolute error of prediction 2.62 years. We do not have an independent validation sample set for semen-DNA based MethAge model so far, so the fair comparison is with the training sets in other publications.

The comparison of our results with a training set by Pisarek et al., MAE = 4.3 years [3] or training set by Xiao et al., MAE = 3.89 years [4] is favorable. Result of training set of the model developed by Jenkins et al. was better than ours (MAE = 2.04 years and 2.62 years, respectively), but it used much bigger array of CpG loci [5].

Within our sample set, non-sperm cell fraction after a differential extraction provided more precise age prediction than sperm fraction, and, surprisingly, also more precise prediction than the whole sperm. The number of sperm cells in ejaculate is highly heterogenous, varying from zero to hundreds of million per milliliter interindividually, as well as varying in one individual in different time and phases of intercourse. Non-sperm cell fraction is probably less variable, consisting mostly of epithelial cells and leukocytes [6], that are commonly used for age prediction.

Contradictory with our results, both models developed by Xiao et al. (two SNaPshot assays containing 11 non-overlapping CpGs each) were more precise for sperm cell than the whole semen samples [4]. This should be attributed mainly to the CpG selection, as Xiao et al. were specifically focused on sperm-specific markers.

Preliminary results of testing samples stored in cryoprotective medium PreservCyt suggest, that in the specific loci used by MethAge model, methylation is not considerably influenced.

The main limitation of our epigenetic age prediction model is the small sample set.

Conclusion

The MethAge method is capable of predicting the age of an unknown semen donor for 89.1% of samples in the error range ± 5 years and a mean absolute error of 2.62 years.

Pilot data from differential extraction suggest that the determining factor for age prediction from semen is not the sperm fraction but the non-sperm fraction. Pilot data also suggest that the cryoprotective medium PreservCyt does not significantly alter age prediction results. More samples will be tested in the future.

BisQuE is a method capable of distinguishing incomplete bisulfite conversion. All tested samples converted in our lab using EZ Methylation Gold kit (Zymo) were converted with 100% efficiency. Yet, due to inconsistent performance and repeated dilution series failure, BisQuE is not the robust method for bisulfitation assessment in our hands.

Acknowledgments

This work was funded by LM2023053 EATRIS-CZ, LM2023033 BBMRI.CZ, LX22NPO5102, TN02000109, and IGA LF UP 2023_006. We want to thank Doctor Sae Rom Hong and Professor Kyoung-Jin Shin for kindly consulting our work.

Conflict of interest statement

Authors declare no conflict of interest.

References

[1] Z. Khosravizadeh, K. Khodamoradi, Z. Rashidi et al., Sperm cryopreservation and DNA methylation: possible implications for ART success and the health of offspring, J Assist Reprod Genet, 39 (2022). https://doi.org/10.1007/s10815-022-02545-6

[2] S.R. Hong and K.-J. Shin, Bisulfite-Converted DNA Quantity Evaluation: A Multiplex Quantitative Real-Time PCR System for Evaluation of Bisulfite Convertion, Front Genet. 12 (2021). doi.org/10.3389/fgene.2021.618955

[3] A. Pisarek, E. Pośpiech, A. Heidegger et al., Epigenetic age prediction in semen – marker selection and model development, Aging 13 (2021). doi.org/10.18632/aging.203399

[4] Ch. Xiao, Y. Li, M. Chen et al., Improved age estimation from semen using sperm-specific age-related CpG markers, Forensic Science International: Genetics, 67 (2023). doi.org/10.1016/j.fsigen.2023.102941

[5] T.G. Jenkins, K.I. Aston, B. Cairns et al., Paternal germ line aging: DNA methylation age prediction from human sperm, BMC Genomics, 19:763 (2018). doi.org/10.1186/s12864-018-5153-4

[6] L. Kotková and J. Drábek, Age-Related Changes in Sperm DNA Methylation and their Forensic and Clinical Implications, Epigenomics 15 (2023). doi.org/10.2217/epi-2023-0307

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 619-623
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Emerging Technologies for Hair Examination and Analysis: Assessing the Profiling Capabilities of Genetic Analysers

Alicia M. Haines[1,2], Tabarek Aljumaili[2,3], Tash Breen[2], David Bruce[2,3], Thomas Jeffries[2]

[1]*School of Mathematical and Physical Sciences, University of Technology Sydney, NSW, Australia.*
[2]*School of Science, Western Sydney University, Hawkesbury, NSW, Australia.*
[3]*NSW Health Pathology, Forensic & Analytical Science Service, Sydney, NSW, Australia.*

** Corresponding author at: Institutional address, Alicia.haines@uts.edu.au*

Abstract

Biological traces play a critical role in DNA casework; however, despite the significance of DNA profiling, its efficacy is often hampered by the limited availability of DNA. Standard DNA analysis has predominantly focused on nuclear DNA and has overlooked other potential biological traces, such as microbiome signatures. Hair samples pose a challenge as they may offer limited nuclear DNA information. This study investigated a novel approach for the analysis of hair samples, integrating traditional nuclear DNA identification with microbial forensic analysis to obtain a comprehensive profile for identification.

This research employed the emerging technologies RapidHIT™ ID for nuclear DNA detection and a SeqStudio® Genetic Analyzer for processing biological traces. Initially, DNA-containing material on hair samples was detected using Diamond™ Nucleic Acid Dye (DD) and real-time Extended Depth of Field (EDF) imaging to visualize and count nuclei. Detection of DNA containing material was an integral step to determine hair suitability for DNA analysis. Hair samples, both shed and plucked, were imaged using the Optico N300F LED Fluorescent Microscope and a MIchrome 5 Pro camera. Subsequently, the samples underwent processing through the RapidHIT™ID platform, recollected and processed through standard DNA analysis.

This study explored microbial signatures in follicle tissue to gauge their discriminatory potential for hairs with limited nuclear DNA and STR profiling results. Bacterial isolates were PCR-amplified and sequenced using the 16S rRNA gene, analysed via MicrobeBridge™ software for forensic relevance. The SeqStudio® platform enabled combined sequencing and STR profiling on a single plate. These findings could significantly impact casework by enhancing hair examination for forensic purposes.

Keywords
Diamond™ Nucleic Acid Dye, Fluorescent Microscopy, Hair Examination

1. Introduction

Biological traces are crucial in DNA casework, but the limited availability of DNA on some sample types often hampers DNA profiling efficacy. Standard DNA analysis has primarily focused on nuclear DNA, overlooking other potential biological traces like microbiome signatures [1]. Hair samples can be challenging due to their limited nuclear DNA information, however, are prevalent at crime scenes due to the high number of hairs a person is likely to shed every day (50-100) [2].

This study used the RapidHIT™ ID (RHID) for nuclear DNA detection and a SeqStudio® Genetic Analyzer for processing other biological traces. DNA-containing material on hair samples can initially be detected using Diamond™ Nucleic Acid Dye (DD) and real-time Extended Depth of Field (EDF) imaging, previous studies have shown that detection of DNA material was crucial to assess hair suitability for DNA analysis [3, 4]. Hair samples, including shed and plucked, were imaged using the Optico N300F LED Fluorescent Microscope and a MIchrome 5 Pro camera. The samples then underwent processing through the RHID platform before being subjected to standard DNA analysis.

In addition to the RHID platform this study explored microbial signatures in follicle tissue to assess their potential for hairs with limited nuclear DNA and STR profiling results. The use of 16S gene is considered the gold standard for identification at species level with numerous established protocols being available [5]. This study evaluated the 16S Direct Workflow [5] and the SeqStudio® platform which can combine sequencing and fragment length analysis (STR profiling) on a single plate.

2. Material studied, methods, techniques

Human Ethics approval number: H15115

2.1 Human identification using RHID workflow

The processing of hair samples through the RHID platform followed the method previously described [3], however, one step further was investigated by extracting the remaining hair sample that was collected post-RHID. Hairs were extracted using the PrepFiler™ BTA kit, followed by quantification using

Quantifiler® Trio half reaction volume and amplified using GlobalFiler™ (GF) Amplification PCR kit (Thermo Fisher Scientific) also at half reaction volume. Amplified products were analysed on a SeqStudio™ Genetic Analyser, following manufacturer's protocol. See Figure 1 for schematic workflow.

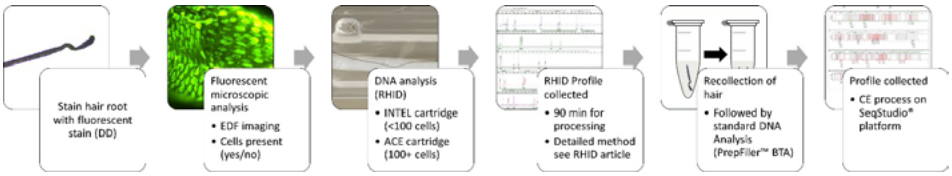


Figure 1. Standard and RHID workflow for the analysis of hair samples

2.2 Microbiome identification using SeqStudio™ workflow

2.2.1 Bacterial isolates: Hair samples freshly plucked from scalp were placed into 10 mL of brain, heart, infusion (BHI) broth or nutrient broth (NB), enriched overnight at 37 °C, then streaked on agar. Isolated colonies were grown overnight at 37 °C in respective broths. Bacterial isolates were extracted using the DNeasy PowerSoil Pro Kit (Qiagen) following manufacturer's protocol.

2.2.2 16S workflow sequencing samples were produced following the 16S Direct workflow specifically for the SeqStudio™ [5], see Figure 2 for schematic workflow. This included using the BigDye™ Direct Cycle sequencing kit (Thermo Fisher Scientific) for the amplification of 16S A and B fragments [5] and then for the sequencing, following the manufacturer's protocol. Sequences were analysed using the MicrobeBridge (Applied Biosystems) software and searched using MicrobeNet a CDC Virtual Reference Laboratory.

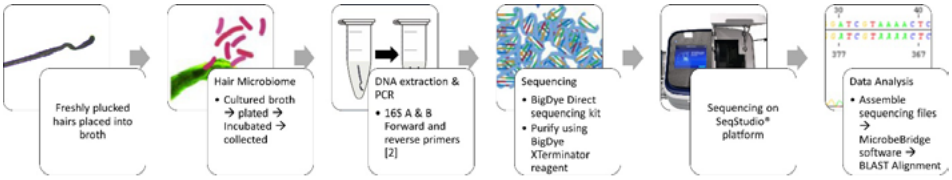


Figure 2. Microbiome workflow for hair analysis

3. Results

3.1 RHID workflow

Hairs that were collected post-RHID analysis were put through standard DNA extraction, quantification, amplification with detection on the SeqStudio™ Genetic Analyser. Hairs were stained with DD and imaged post-RHID, a noticeable amount of cellular material was removed during the RHID process. RHID uses a crude extraction process, which results in hairs not being consumed during the process and can be recollected for standard DNA analysis [3]. Average allele percentage recovered using RHID for shed hairs was ~4.5% which increased to 17% when processed through standard DNA analysis. This may in part be due to the full breakdown of the hair sample during the standard DNA analysis process. DNA profiles were successfully obtained from these hair samples that were recollected, with concordant results.

3.2 Microbiome identification

There was a 70.8% fail rate of the sequencing runs (48 total runs) through the SeqStudio™ platform. Either due to, no sample being detected or basecalling failed due to poor quality data. Only partial sequencing results were able to be obtained due to the fail rate. However, the partial sequencing result was able to be analysed using the MicrobeBridge software, the aligned sequence was then searched through the BLAST function in MicrobeNet and was identified as Staphylococcus.

4. Discussion

4.1 RHID and Standard DNA analysis:

The recollection of hair samples from the RHID cartridge and processed through standard DNA analysis was successful in this study. Full profiles were obtained from plucked hairs that were concordant with the RHID results.

4.2 Microbiome identification:

Further method optimization is required, the protocol outlined in the 16S Direct workflow [5], was quite labour intensive and expensive. For this workflow to be useable in a casework environment within forensic science, there needs to be a higher reliability of the protocol and success in the results.

5. Conclusion

Standard DNA analysis on hairs collected post-RHID generally resulted in improved DNA profiling results, compared to the RHID profiles. Demonstrating the ability to obtain a DNA profile through standard DNA workflow.

Further investigation is required on whether 16S sequencing using the SeqStudio™ Platform and BigDye™ Direct chemistry, is possible from hair samples extracted for Human Identification purposes, not just bacterial isolates.

6. Acknowledgements

Tabarek Aljumaili for the work completed during her Master of Research and WSU for the funding provided through the SoSc Collaborative Grant Scheme.

7. Conflict of interest statement

No conflict of interest.

8. References

[1] A.A. Singh, and M. Okpeku, Emerging methods of human microbiome analysis and its forensic applications: Review. *FSI: Reports*, 2024. **9**: p. 100355.

[2] J.R. Robertson, Forensic Examination of Hair. 1 ed. Taylor & Francis forensic science series. 1999, London: CRC Press LLC.

[3] T. Aljumaili, and A.M. Haines, An evaluation of the RapidHIT™ ID system for hair roots stained with Diamond™ Nucleic Acid Dye*, FSI: Genetics*, 2024.

[4] A.M. Haines, and A. Linacre, *A rapid screening method using DNA binding dyes to determine whether hair follicles have sufficient DNA for successful profiling.* Forensic Science International, 2016. **262**: p. 190-195.

[5] Application Note, The 16S Direct workflow - Microbial identification using 16S gene sequencing on the SeqStudio Genetic Analyzer and analysis with MicrobeBridge software, *Thermo Fisher Scientific*.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 625-631
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Performance Comparison of Two RapidHIT™ ID Cartridge Types for Forensic Leads

Hong Han LIM[1*], Baoqiang HENG[1], Christopher Kiu-Choong SYN[1]

[1]*DNA Profiling Laboratory, Health Sciences Authority, 11 Outram Road, S169078, Singapore*

** Corresponding author at: 11 Outram Road, S169078, Singapore, LIM_Hong_Han@hsa.gov.sg.*

Abstract

The RapidHIT™ ID is an automated system which can generate a DNA profile in 90 minutes from sample insertion. This study compares the performance of the newly released RapidINTEL™ Plus (INTEL+) cartridge against the ACE GlobalFiler™ Express (ACE) cartridge. Both cartridge types yielded full profiles from buccal swabs. However, peak heights of samples processed with INTEL+ were generally two times higher, suggesting greater sensitivity. This study also assessed the analysis of swabs retrieved from expended cartridges. Full profiles were recovered when buccal swabs retrieved from both cartridge types were reprocessed with the standard laboratory workflow. We also evaluated a hitherto unknown approach of retrieving waste pads from expended cartridges for DNA analysis and full profiles were successfully recovered. Interestingly, quantitation data showed that the DNA recovered from the retrieved swabs was approximately 8 to 10 times more than the DNA recovered from the waste pads, which indicated that the extraction efficiency of RapidHIT was limited. In conclusion, our study has demonstrated the improved sensitivity of the INTEL+ cartridge, and also demonstrated two possible approaches of repeat analysis using either swabs or waste pads retrieved from expended cartridges.

Keywords

ACE GlobalFiler™ Express, RapidINTEL™ Plus, buccal swab.

Introduction

The standard laboratory workflow for DNA analysis involves multiple hands-on procedures and generally takes at least 6 hours to generate a DNA profile. In contrast, the RapidHIT™ ID system is an automated instrument that requires minimal human intervention and only 90 minutes to generate a DNA profile from

sample insertion [1-3]. As this system can only process one sample at a time, its main use would be in the analysis of a limited number of urgent, critical samples for investigators to receive investigative leads and detain suspects based on DNA evidence. The RapidHIT™ ID system utilizes two types of sample cartridges, which are (i) the RapidHIT™ ID ACE GlobalFiler™ Express (ACE) cartridge meant for high DNA samples, and (ii) the recently launched RapidINTEL™ Plus (INTEL+) cartridge with higher sensitivity, capable of analyzing lower amounts of DNA [4]. Data on the performance of the INTEL+ cartridge is currently limited.

It was reported previously that buccal swabs retrieved from expended sample cartridges can be processed via the standard laboratory workflow to obtain full profiles [3]. Other than the retrieved swabs, we theorized that waste pads from the waste chamber in expended cartridges could also be an alternative source of DNA for analysis. This is because during RapidHIT analysis, 500 µL of lysate is produced in the sample chamber [4]. However, the volume of the PCR chamber is only about 12 µL [1]. As such, the remaining lysate will have to be moved to the waste chamber, which makes the waste pad a possible source of DNA.

Our study therefore aims to (i) perform a comparison on the performance of ACE and INTEL+ cartridges using buccal swabs, and (ii) assess the potential of recovering DNA from swabs and waste pads retrieved from expended cartridges.

Material studied, methods, techniques

Sample collection

Buccal cells were collected from three volunteers by performing 5 swipes on the buccal surface using cotton swabs.

Waste pads were excised from waste chambers of expended sample cartridges using scalpels.

RapidHIT™ ID workflow

Buccal swabs were processed with either ACE (n=6) or INTEL+ (n=6) with General Protocol selected, on the RapidHIT™ ID system (Thermo Fisher Scientific, USA) with RapidLINK™ v2.0 software. Electrophoresis data was analyzed with GeneMarker® HID v2.95.

Standard laboratory workflow

Buccal swabs and waste pads (measuring ~2.5 cm × 1 cm) retrieved from expended cartridges were processed in whole as follows: DNA extraction was performed

with the DNA IQ™ kit on the Maxwell® FSC instrument (Promega Corporation, USA) with a constant DNA elution volume of 56 µL. Quantitation was performed with the Quantifiler™ Trio Kit on the QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystems, USA). Amplification was performed with 1 ng of DNA template using the GlobalFiler™ Amplification Kit. DNA separation and detection were performed using the 3500xL Genetic Analyzer (Applied Biosystems, USA), with injection parameters of 3 µL DNA, 1.2 kV, 24 s. Electrophoresis data was analyzed using GeneMapper™ ID-X v1.6.

Data analysis

For the analysis of allele peak heights, the height of each homozygous peak was divided by 2. Heterozygote peak height ratio was calculated by dividing the height of the shorter peak by the height of the taller peak in a heterozygous locus.

Results

Allele recovery: *To compare the performance of the ACE and INTEL+ cartridges, buccal swabs were analyzed with both cartridges. Our results showed that RapidHIT analysis using both cartridge types achieved full allele recovery from buccal swabs. Subsequently, the swabs were retrieved from the cartridges and were processed via the standard laboratory workflow. We achieved full allele recovery from all swabs retrieved from the expended ACE and INTEL+ cartridges. We also investigated the feasibility of recovering alleles from waste pads excised from expended cartridges using the standard laboratory workflow. Our results showed that full allele recovery was achieved from the waste pads, with the exception of two waste pads retrieved from INTEL+ cartridges where there was one allelic dropout in each waste pad profile. Using buccal swabs, we demonstrated that full profiles can be generated from swab reprocessing and waste pad processing via the standard laboratory workflow.*

We also compared the DNA yield from swabs and waste pads retrieved from expended cartridges. Our quantitation data showed that the mean DNA concentration in swabs retrieved from ACE and from INTEL+ were 9.12 ±2.04 ng/µL and 8.03 ±1.70 ng/µL, respectively. The mean DNA concentration in waste pads retrieved from ACE and from INTEL+ were 1.09 ±0.73 ng/µL and 0.76 ±0.28 ng/µL, respectively. The amount of DNA in INTEL+ waste pads was marginally lower than that in ACE waste pads. This observation is consistent with INTEL+ having an additional

paper punch in the PCR chamber for lysate capture, thereby resulting in less lysate entering the waste chamber.

Allele peak heights: *Next, allele peak heights generated from the ACE and INTEL+ cartridges were compared. Comparison of allele peak heights showed that buccal swabs processed with ACE and INTEL+ yielded a mean p*eak height of 756 ±422 RFU and 1,702 ±730 RFU, respectively (Fig. 1). Swabs and waste pads retrieved from expended cartridges processed via the standard laboratory workflow yielded similar peak heights. No comparison was made between peak heights from RapidHIT analysis against peak heights from reprocessed swabs or waste pads due to inherent differences in PCR chemistry and detection system of the two workflows.

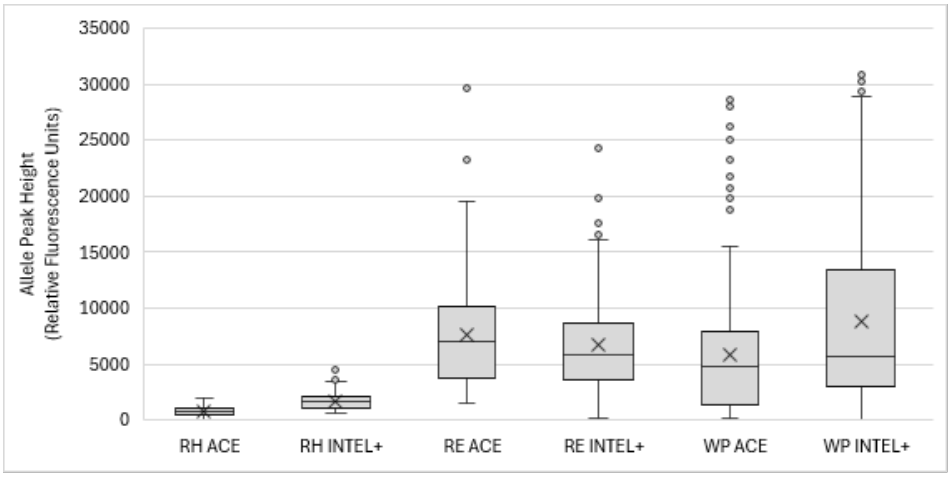


Figure 1. Distribution of peak height of every allele from RapidHIT processing, swab reprocessing via standard laboratory workflow, and waste pad processing via standard laboratory workflow. N=6 for each box. Horizontal line in box indicates the median. Bottom of box indicates first quartile. Top of box indicates third quartile. Extremes of whiskers indicate minimum and maximum values. Cross in box indicates the mean. Circles indicate outliers. RH ACE: Buccal swabs processed with ACE cartridge, RH INTEL+: Buccal swabs processed with INTEL+ cartridge, RE ACE: Reprocessed swabs from ACE cartridge, RE INTEL+: Reprocessed swabs from INTEL+ cartridge, WP ACE: Waste pads from ACE cartridge, WP INTEL+: Waste pads from INTEL+ cartridge.

Heterozygote peak height balance: Similar peak height ratios were obtained from buccal swabs analyzed using ACE and INTEL+, *with most values above 0.8* **(Fig.**

2*). This result was also observed from the buccal swabs as well as the waste pads retrieved from ACE and INTEL+ and processed with the standard laboratory workflow.*

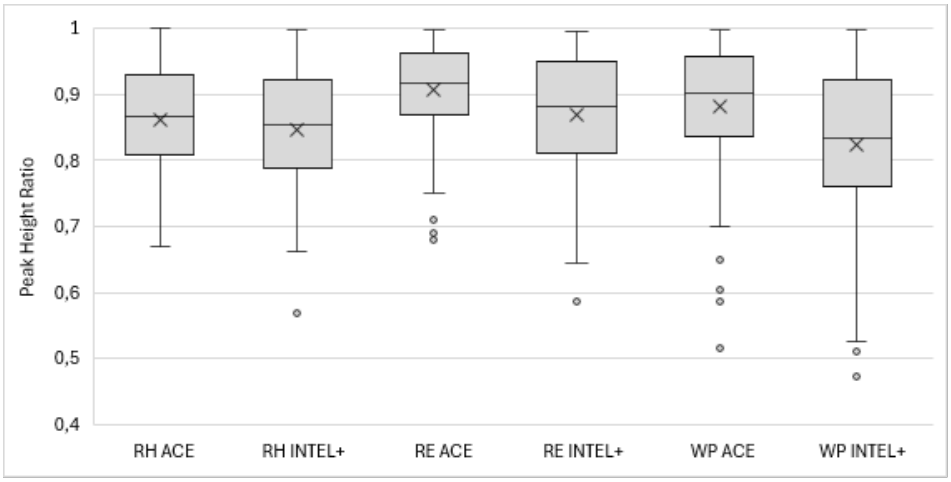


Figure 2. Distribution of peak height ratio of every heterozygous locus from RapidHIT processing, swab reprocessing via standard laboratory workflow, and waste pad processing via standard laboratory workflow. N=6 for each box. Horizontal line in box indicates the median. Bottom of box indicates first quartile. Top of box indicates third quartile. Extremes of whiskers indicate minimum and maximum values. Cross in box indicates the mean. Circles indicate outliers. RH ACE: Buccal swabs processed with ACE cartridge, RH INTEL+: Buccal swabs processed with INTEL+ cartridge, RE ACE: Reprocessed swabs from ACE cartridge, RE INTEL+: Reprocessed swabs from INTEL+ cartridge, WP ACE: Waste pads from ACE cartridge, WP INTEL+: Waste pads from INTEL+ cartridge.

Discussion

Our comparison of the performance of the ACE and INTEL+ cartridges showed that the mean peak height from buccal swabs processed with INTEL+ was more than twice of that from using ACE. This observation is likely due to (i) INTEL+ cartridge having an additional paper punch for lysate capture in the PCR chamber, and (ii) having two additional PCR cycles in the INTEL+ protocol [4]. In this respect, the INTEL+ cartridge may be the cartridge of choice when processing samples of lower DNA quantity.

Our study demonstrated that swabs retrieved from expended ACE and INTEL+ cartridges can be reprocessed to yield high quality profiles, corroborating results reported by Cihlar et al. [3]. Laurin et al. had previously suggested that the efficiency of RapidHIT was limited, as samples processed by RapidHIT and their unprocessed controls yielded comparable DNA quantities [5]. Our quantitation data showing that retrieved swabs yielded about 8 to 10 times higher DNA concentration than waste pads supported this conclusion. Given this finding, laboratories have the option of using RapidHIT for quick DNA leads. The same swab can subsequently be retrieved for reprocessing via standard laboratory workflows for official reporting.

Our study also provided the first report (to our knowledge) that DNA profiles can be obtained by processing waste pads via the standard laboratory workflow. This approach confers two main advantages. First, the waste pad may be the only sample source available when the original sample in the sample chamber disintegrates during RapidHIT analysis. Second, the expended cartridge can be readily stored at 4°C, and the waste pad remains preserved in a sealed, enclosed chamber thereby ensuring sample integrity. The latter may be of particular relevance if the results are challenged and independent result verification is required by, for example, defense experts as any sign of cartridge tampering would be clearly visible.

Conclusion

Our study demonstrated that RapidHIT analysis of buccal swabs using INTEL+ yielded approximately twice the peak heights when compared to using ACE. This result indicates that INTEL+ may be the cartridge of choice over ACE when low-DNA samples need to be analyzed. We also showed that using the standard workflow, full profiles can be obtained from swabs and waste pads retrieved from expended cartridges, making them viable alternatives in the event that result confirmation is needed.

Acknowledgments

We express our appreciation to all staff of the DNA Profiling Laboratory for their support throughout the study.

Conflict of interest statement

None.

References

[1] Salceda S, Barican A, Buscaino J, Goldman B, Klevenberg J, Kuhn M, Lehto D, Lin F, Nguyen P, Park C, Pearson F, Pittaro R, Salodkar S, Schueren R, Smith C, Troup C, Tsou D, Vangbo M, Wunderle J, King D. Validation of a rapid DNA process with the RapidHIT® ID system using GlobalFiler® Express chemistry, a platform optimized for decentralized testing environments. Forensic Sci Int Genet. 2017 May;28:21-34. doi.org/10.1016/j.fsigen.2017.01.005. Epub 2017 Jan 12. PMID: 28135583.

[2] Murakami C, Irie W, Sasaki C, Nakamaru N, Sakamoto M, Nagato J, Satoh F. Individual identification using the RapidHIT™ ID system for forensic samples. Leg Med (Tokyo). 2020 Nov;47:101776. doi.org/10.1016/j.legalmed.2020.101776. Epub 2020 Aug 6. PMID: 32795931.

[3] Cihlar JC, Kapema KB, Budowle B. Validation of the Applied Biosystems RapidHIT ID instrument and ACE GlobalFiler Express sample cartridge. Int J Legal Med. 2022 Jan;136(1):13-41. doi.org/10.1007/s00414-021-02722-9. Epub 2021 Oct 13. PMID: 34643802; PMCID: PMC8511620.

[4] Thermo Fisher Scientific. RapidINTEL™ Plus Sample Cartridge User Bulletin. 2024.

[5] Laurin N, Boulianne H, Frégeau C. Comparative analysis of two Rapid DNA technologies for the processing of blood and saliva-based samples. Forensic Sci Int Genet. 2023 Nov;67:102928. doi.org/10.1016/j.fsigen.2023.102928. Epub 2023 Aug 8. PMID: 37573630.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 633-640
DOI: https://dx.doi.org/10.15304/cc.2025.1869

STR genotyping by real-time PCR using QueSTR probes

Sonja Škevin[1,2], Olivier Tytgat[1,2], Maarten Fauvart[2], Liesl De Keyzer[1], Dieter Deforce[1]
Filip Van Nieuwerburgh[1,*]

[1]*Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium*
[2]*imec, Leuwen, 3001, Belgium.*

** Corresponding author at: Ottergemsesteenweg 460, Gent, 9000, Belgium,*
Filip.VanNieuwerburgh@UGent.be.

Abstract
QueSTR probes provide a feasible alternative for implementing forensic Short Tandem Repeat (STR) genotyping on lab-on-a-chip (LoC) devices. Unlike capillary electrophoresis (CE), which poses challenges for LoC implementation, QueSTR probes have been demonstrated to accurately genotype STRs in a hybridization curve assay after PCR amplification. In this study, we modified the QueSTR probe assay for use as a hydrolysis probe assay during qPCR. An asymmetric real-time PCR was performed with QueSTR probes and RNase H2 in the master mix, during which the fluorescence was recorded. Designed to hybridize with specific STR alleles, QueSTR probes con-tain a fluorescent dye and quencher molecule, enabling detection of probe hybridization through RNase H2-mediated cleavage of the RNA:DNA duplex, which releases the quencher. Matching probes yielded lower threshold cycle values and steeper incline of fluorescence curves compared to non-matching probes, indicating accurate genotype. The QueSTR qPCR assay was used to successfully genotype four CODIS core loci (D16S539, D7S820, TPOX, and TH01) in 12 samples, with one exception. The integration of amplification and detection in a single reaction supports the use of QueSTR probes for miniaturizing STR genotyping, thereby complementing CE-based analyses in centralized labs.

Keywords
Short tandem repeats, forensic genetics, QueSTR probes, real-time PCR

Introduction

Short tandem repeats (STRs) are genomic regions where 2-6 nucleotides are repeated in a head-to-tail fashion. The number of repeats at a given locus varies among individuals, making them useful for kinship analysis and identification of individuals (1). Typically, STR profiling involves PCR amplicon sizing via capillary electrophoresis (CE) or next-generation sequencing, which requires bulky equipment and trained personnel, resulting in high upfront costs (2). As a result, there is a growing demand for portable STR genotyping devices for use in the field, particularly in police stations, crime scenes and mass disasters. Lab-on-a-chip (LoC) technology has the potential to overcome the limitations of the conventional STR genotyping process, while providing rapid and accurate information on-site (3). Small reaction volumes of LoCs allow faster thermocycling, resulting in a faster analysis as well as lower reagent consumption (4)(5). Several probe-based assays have been developed as an approach for the STR genotyping assay miniaturization and application a LoC, two of which were developed by our research group: STRide probe melting curve assay (6) and the QueSTR probe hybridization curve assay (7). In the current study, the QueSTR probe assay is modified as a hydrolysis probe assay during a qPCR reaction instead of in a hybridization curve analysis after PCR, resulting in a much-simplified process.

The QueSTR probes (Figure 1) are synthetically manufactured oligonucleotides consisting of three regions: an anchor region (blue), a repeat region (yellow-orange), and a sensor region (green). The repeat region is modified with a 6-carboxyfluorescein (FAM) fluorophore, while the sensor region is modified with a quencher. The sensor region also contains an RNA moiety, separating the fluorophore and the quencher (7). The QueSTR probe assay utilizes the ability of RNase H2 to cleave the RNA:DNA duplex, which is formed when the sensor region hybridizes with the sample. This releases the quencher and allows for fluorescence to be emitted. This study introduces a new method for real-time PCR genotyping of forensic STRs, in which the QueSTR probes and RNase H2 are added to the PCR master mix, and the QueSTR probes are used as hydrolysis probes. This approach offers greater convenience for LoC applications than the previously published QueSTR hybridization curve assay (7), allowing amplification and detection in a single reaction in a single cavity. We describe the implementation of QueSTR probes in qPCR for forensic STR genotyping of four CODIS core loci with a relatively low number of common alleles (TH01, D7S820, D16S539, and TPOX).

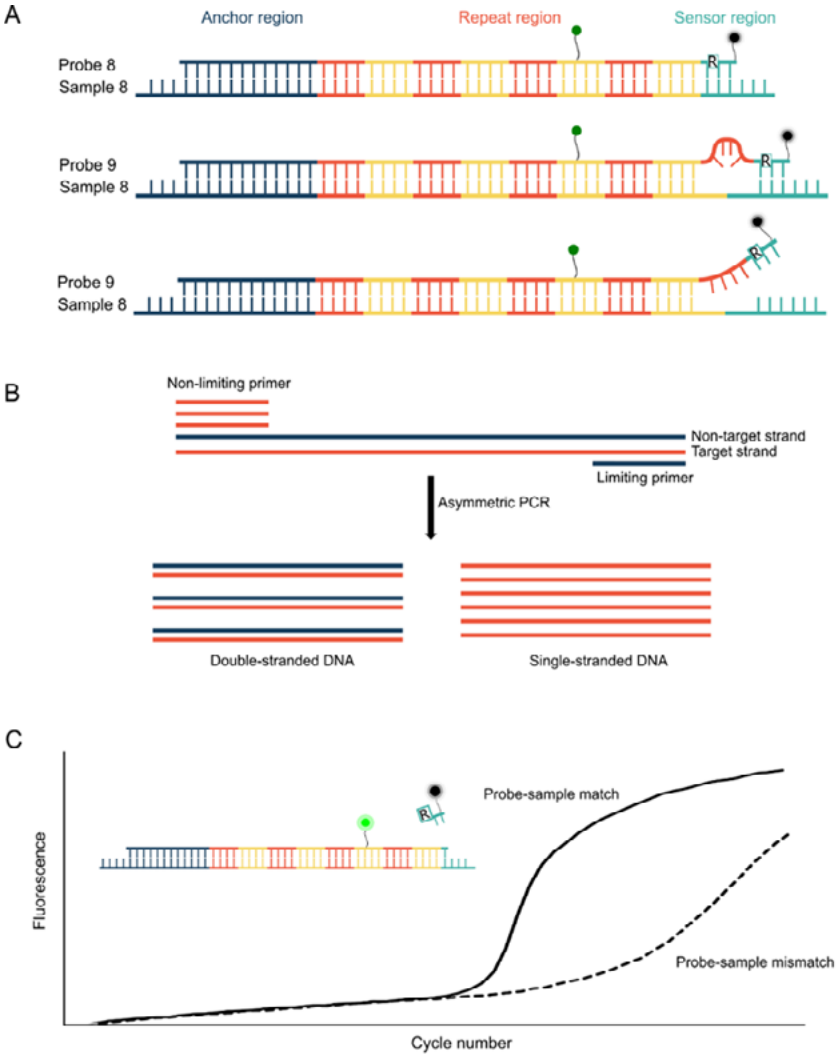


Figure 1. A) QueSTR probe design. *The probes consist of three regions: an anchor region (blue), a repeat region (yellow-orange), and a sensor region (green). The sensor is separated from the repeat region by one RNA moiety. The repeat region is labelled with a FAM-fluorophore, while the quencher is attached to the terminus of a sensor region.* **B) Asymmetric amplification.** *Primer that is priming the strand complementary to the probe is added in excess, which results in single-stranded PCR product complementary to the probe.* **C) qPCR.** *Probes are added to the master mix of the reaction, and fluorescence is recorded in real-time during the amplification. Sigmoidal curves of the matching probes appear at the lower Ct and have steeper incline.*

Material studied, methods, techniques

Sample collection and DNA extraction

Four loci were tested on a total of 12 samples: two commercially available reference DNA samples (9947a, 9948) obtained from Promega (Madison, WI, USA), five DNA samples extracted from saliva (samples 1-5), and five DNA samples extracted from blood (samples 6-10). DNA was extracted from saliva by simple DNA extraction using chelex resin purchased from Bio-Rad (Bio-Rad Laboratories, CA, USA) by the protocol outlined by Sweet et al. (8)preservation and analysis of body fluid stains is an important aspect of forensic science, PCR-based typing of DNA extracted from recovered stains is often a crucial method to identify a perpetrator or exclude an innocent suspect. This paper reports an improved method of extracting genomic DNA from saliva stains deposited on human skin in simulated bite mark situations. Results of organic (phenol-chloroform with an input of 1000 µL of whole saliva. The extraction of DNA from the blood samples was performed following the manufacturer's protocol, using the DNeasy® Blood and Tissue kit (Qiagen, Helden, Germany).

CE genotyping

CE was used as a reference genotyping method for the collected samples. Extracted DNA was amplified by PCR using the AmpFlSTR® Identifiler® Plus PCR amplification kit (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's protocol. CE was performed using the ABI3130xl Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA, USA). Obtained electropherograms were analyzed using the GeneMapper ID-x 1.2 software (Thermo Fisher Scientific, Waltham, MA, USA).

Asymmetric qPCR probe assay

An asymmetric locus-specific qPCR starting from 1 ng of DNA input in a total reaction volume of 10 µl was performed. A lower concentration of one primer was used compared to the other, resulting in an excess of a DNA strand complementary to the probe. The master mix comprised of KAPA2G reaction buffer at 1X (Roche, Basel, Switzerland), MgCl2 at 1.5 mM, dNTPs (Thermo Fisher Scientific, Waltham, MA, USA) at 0.2 mM each, and 1 U of KAPA2G Fast HotStart polymerase (Roche, Basel, Switzerland). For all loci, in each reaction, the excess primer was present at a concentration of 1.5 µM. The limiting primer was present at a concentration of 0.1 µM for locus TH01 and 0.2 µM for loci D7S820, D16S539, and TPOX. RNase

H2 (Integrated DNA Technologies, Newark, NJ, USA) and allele-specific QueSTR probes were added directly to the master mix at concentrations of 0.005 U/µL and 0.1 µM, respectively. Temperature cycling consisted of an initial denaturation step of 2 minutes at 95 °C followed by 60 cycles consisting of a denaturation step for 30 s at 95 °C followed by 60 s of annealing/elongation at respective temperatures for each locus. Fluorescence was recorded at the annealing/elongation step of each cycle. The fluorescence curve with the lowest Ct value, steepest incline, and highest plateau fluorescence was called as corresponding to an allele being present in the sample. If the difference between the curve with the lowest Ct value and the following ones was large, the sample was called homozygous for the allele corresponding to the curve with the lowest Ct value.

Results

Accuracy and replicability assessment

Genotyping accuracy was assessed for all four loci (D7S820, D16S539, TH01, and TPOX) by testing 12 DNA samples in triplicate, including 5 DNA samples isolated from blood, 5 DNA samples isolated from saliva, and 2 commercially available forensic reference samples. All samples were correctly genotyped across all four loci, except for a sample with true genotype 8:8 for locus D7S820.

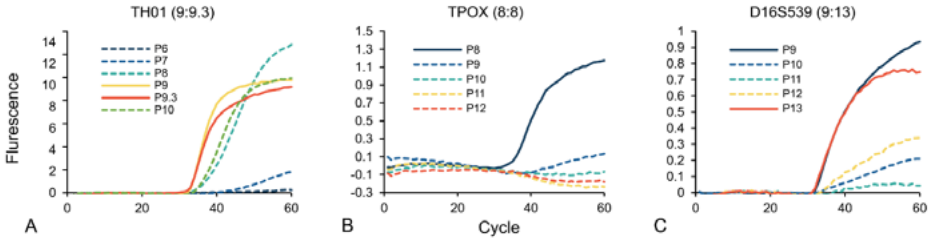


Figure 2. qPCR graphs for three different samples. Locus and true genotype are indicated in the chart title. Fluorescence curves of matching probes are indicated in full lines, while mismatch curves are represented as dashed lines.

Sensitivity assessment

To test the sensitivity of the assay, one sample for each locus was serially diluted, resulting in inputs ranging from 31 pg to 500 pg. Correct genotypes could be

called for the loci D16S539, TH01, TPOX at 31 pg, equivalent to the DNA content of 6 cells. The correct genotype for D7S820 locus was established at 62 pg of input DNA, while an input of 31 pg resulted in a dropout

Discussion

Figure 2 shows representative results for alleles and allele combinations that can be expected to be challenging in probe-based STR genotyping. Figure 2A shows an example of a sample with true TH01 genotypes 9 and 9.3. The exponential phase of probe 9 and 9.3 curves starts at virtually the same Ct, while the Ct value of the probe 10 curve, which is only one nucleotide longer, is noticeably higher. The incline of the probe 10 curve is also less steep than for the curves corresponding to the true alleles. Figure 2B shows an example of a sample with a homozygous TPOX allele 8:8 genotype. Theoretically, the homozygous samples are challenging to call due to stutter amplification. However, homozygous samples were called correctly in all cases except for the 8:8 genotype for the D7S820 locus. Figure 2C shows an example of a sample with true D16S539 genotypes 9 and 13. Samples containing two alleles with a large repeat number difference between them are challenging due to the significant difference in probe hybridization temperatures caused by the difference in length between them. We were able to genotype such samples correctly, an example of which is shown in Figure 2C.

All loci were genotyped correctly with and input of 31 pg of DNA except D7S820 which was genotyped correctly at 62 pg of DNA input. This difference in sensitivity can probably be attributed to differences in PCR efficiency between loci. The relatively high annealing temperature used in this assay requires longer primers. These primer lengths can cause issues such as the formation of secondary structures and primer dimers that may impact PCR efficiency. Nevertheless, the assay demonstrated high sensitivity, making it suitable for applications with only a limited amount of input DNA.

Conclusion

In this study, the previously published QueSTR probes were used in a qPCR assay. The probes are labelled with a fluorophore and a quencher, separated from each other by an RNA moiety. The assay utilizes RNase H2 to cleave DNA:RNA duplexes,

resulting in fluorescence emission from the fluorophore. A match between the probe and the sample resulted in fluorescence curves with the lowest Ct values and the steepest inclines. All except one of 12 samples were correctly genotyped across four loci, including challenging genotypes of nearest repeat neighbours, homozygous samples, and samples with large differences in repeat number between the present alleles. The assay showed high replicability and high sensitivity, with the ability to correctly genotype samples with input as low as 31 pg for TH01, TPOX and D16S539 loci and 62 pg for D7S820 locus. While the proof-of-concept study demonstrated the accuracy and sensitivity of the QueSTR probes qPCR assay, further optimization is required to maximize its potential. This could involve refining the probe anchor length, primer length, and fine-tuning the annealing/elongation temperatures. In addition, an automated data interpretation method should be developed for automated, straightforward genotyping

Conflict of interest statement

Sonja Škevin reports financial support in terms of PhD grant that was provided by Research Foundation Flanders [1SC6522N].

Filip Van Nieuwerburgh, Dieter Deforce and Olivier Tytgat are inventors on a pending patent application disclosing the STR-probes described in this publication (#WO 2021/175762 A1).

References

[1] Murat P, Guilbaud G, Sale JE. DNA polymerase stalling at structured DNA constrains the expansion of short tandem repeats. Genome Biol [Internet]. 2020 Aug 21 [cited 2023 Mar 6];21(1):1–26. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02124-x

[2] Butler JM. Advanced topics in forensic DNA typing : methodology. 2012 [cited 2023 Mar 6]; Available from: https://books.google.com/books/about/Advanced_Topics_in_Forensic_DNA_Typing_M.html?hl=nl&id=44tX9Ole490C

[3] Bruijns B, van Asten A, Tiggelaar R, Gardeniers H. Microfluidic Devices for Forensic DNA Analysis: A Review. Biosens 2016, Vol 6, Page 41 [Internet]. 2016 Aug 5 [cited 2023 Mar 6];6(3):41. Available from: https://www.mdpi.com/2079-6374/6/3/41/htm

[4] Weiping Y, Liqun D, Jing W, Lingzhi M, Jianbo Z. Simulation and experimental study of PCR chip based on silicon. Sensors Actuators, B Chem. 2005 Jul 22;108(1-2 SPEC. ISS.):695–9.

[5] Daniel JH, Iqbal S, Millington RB, Moore DF, Lowe CR, Leslie DL, et al. Silicon microchambers for DNA amplification. Sensors Actuators, A Phys. 1998 Nov 1;71(1–2):81–8.

[6] Tytgat O, Fauvart M, Stakenborg T, Deforce D, Van Nieuwerburgh F. STRide probes: Single-labeled short tandem repeat identification probes. Biosens Bioelectron [Internet]. 2021 May 15 [cited 2023 Mar 6];180. Available from: https://pubmed.ncbi.nlm.nih.gov/33690100/

[7] Tytgat O, Škevin S, Fauvart M, Stakenborg T, Deforce D, Van Nieuwerburgh F. QueSTR probes: Quencher-labeled RNase H2-dependent probes for Short Tandem Repeat genotyping. Sensors Actuators B Chem. 2022 Jun 15;361:131714.

[8] Sweet D, Lorente M, Valenzuela A, Lorente JA, Alvarez JC. Increasing DNA extraction yield from saliva stains with a modified Chelex method. Forensic Sci Int. 1996 Dec 27;83(3):167–77.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 641-648
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Relevant sampling areas on firearms for the reconstruction of shooting scenario involving two individuals

Martina Onofri[1,2*], Federica Tommolini[1], Simona Severini[1], Massimo Lancia[2], Eugenia Carnevali[1]

[1]*Forensic Sciences Laboratory, Section of Legal Medicine, Department of Medicine and Surgery, Santa Maria Hospital, University of Perugia, 05100 Terni, Italy.*

[2]*Section of Legal Medicine, Department of Medicine and Surgery, University of Perugia, 06123 Perugia, Italy.*

** Corresponding author at: University of Perugia – Forensic Sciences Laboratory, via C. Mazzieri, 3, Terni, Italy. martina.onofri@live.it*

Abstract

DNA traces recovered from firearms can provide valuable information for identifying individuals involved in firearm-related events. This study consisted of the experimental reproduction of a real shooting incident during which a private semi-automatic handgun was discharged by its owner and a second, unrelated individual. The objective was to determine the most informative areas for identifying the last user of a handgun, provided that they did not own the firearm nor load it prior to firing. Contact traces were collected from the surface of seven private semi-automatic handguns at ten sampling locations in two experimental phases. Phase A investigated the amount and composition of traces recovered from guns handled solely by their owners. In Phase B, the shooting scenario was simulated. The traces' quantitative and qualitative characteristics were evaluated per each sampling area. The last individual to discharge a semi-automatic handgun was successfully retrieved from the firearm's body. The greatest quantity of DNA and the most probative STR profiles for the last shooter were recovered from the grip panels. Scarce and less informative traces were obtained from the trigger. Nonetheless, the owner was consistently identified as a major contributor, and despite the weapon's exclusive ownership and use, non-prevalent DNA was frequently recovered. In conclusion, the characteristics of traces recovered from firearms appear to be predominantly determined by the deposition of DNA over time rather than by the last use of the item.

Keywords

Firearms; Personal Identification; Shooting; Touch DNA

1. Introduction

DNA traces recovered from firearms can provide valuable information for identifying individuals involved in firearm-related events. However, there is a paucity of empirical data on the transfer, prevalence, persistence, and recovery of touch DNA traces from firearms under realistic case-work conditions [1-3]. To evaluate touch DNA transfer, the circumstances of a real shooting incident involving two individuals were reproduced in an experimental setting.

The case: During a physical altercation between two individuals, one of them loaded his semi-automatic handgun and discharged it multiple times. The second individual then allegedly took possession of the weapon and fired two shots at the owner. The gun was dropped to the ground, where it was recovered. Our laboratory was tasked with analysing the firearm to reconstruct the event dynamic and assess whether the second individual could have been the last person to use the gun. Partial profiles attributable to the last shooter were recovered only from the gun's body while the gun owner's DNA was retrieved in substantial amounts from all sampled areas.

The study aimed to determine which sections of a handgun's surface could yield probative results for identifying the last individual to discharge it, provided that they did not own nor load the weapon.

2. Materials and methods

The study was approved by the University of Perugia Ethics Committee (n. 1051062/2022). Seven expert shooters participated in the study and provided their private semi-automatic handguns. The surface of each handgun was divided into ten sampling areas (**Figure 1**).

Figure 1. Sampling areas. A) Back strap; B) Front strap; C) Right and left grip panels; D) External trigger guard; E) Internal trigger guard; F) Trigger, G) Right slide; H) Left slide; I) Upper surface of the slide; L) Magazine.

The traces were collected in two phases:

- **Phase A** investigated the amount and composition of traces recovered from guns handled solely by their owners (O). Reference samples of the participants were also collected after they provided informed consent.
- A two-month period followed in which the owners were asked to use their guns according to their habits to allow for self DNA deposition. To preserve realistic case-work conditions, no restrictions were imposed on the participants.
- **Phase B** took place at a nationally accredited shooting range and simulated the shooting scenario. O loaded the gun and discharged it five times. For safety reasons, the physical altercation was not re-enacted and O placed the loaded gun on a sterile piece of paper. A second individual, labelled as Occasional Shooter (OS), picked up the gun and fired two shots.

Contact traces were collected by double-swabbing the area with sterile cotton swabs (MEUS srl, Piove di Sacco, Italy). An additional collection step was carried

out for knurled surfaces by tape-lifting the area with previously UV-sterilised clear Scotch® Tape. Samples were extracted using the phenol/chloroform organic method [4] and resuspended in 30 µL of sterile water. Extracts were quantified using the PowerQuant® System (Promega) and typed in replicates using the PowerPlex® ESX17 Fast System (Promega). CE was performed on the SeqStudio™ Genetic Analyzer (Thermo Fisher Scientific) and data was collected with the GeneMapper™ ID-X software (v 1.6, Thermo Fisher Scientific).

Samples with 10 or more consolidated loci were deemed suitable for comparison (in accordance with Ge.F.I. guidelines [5]) and were thus statistically evaluated using EuroForMix [6]. Otherwise, manual calculations were performed. The qualitative and quantitative characteristics of the traces were evaluated, including DNA yield, relative mixture proportion, the strength of the evidence (Likelihood Ratio, LR), and profile completeness.

3. Results

In general, a lower DNA recovery was observed for Phase B samples, which is likely due to O's use of the gun in the two-month interval being insufficient to re-establish the amount of DNA deposited prior to Phase A. The results are illustrated in Figure 2.

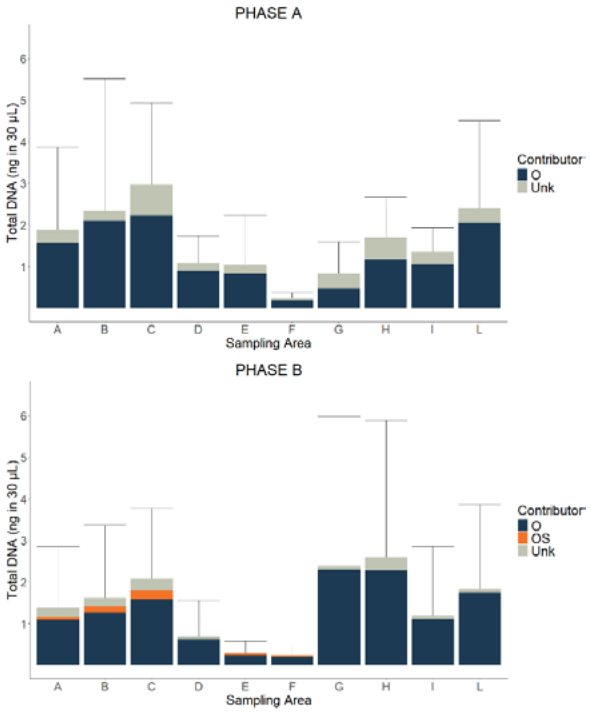


Figure 2. The height of the bar plots falls at the mean DNA amount recovered from each sampling location, the error bar represents the standard deviation, and the relative contribution to the trace is illustrated through colour-coded partitioning of the bar plot.

— **Phase A**: The highest total DNA amount was recovered from sampling areas C, L, B, and A, while the lowest quantities were obtained from area F. Despite the participants being the sole owners and users of the guns, 64% of the traces presented one or more unknown contributors. The partner of one participant was identified as a recurrent contributor to the participant's gun.

— **Phase B**: The highest total DNA amount was recovered from sampling areas H, G, C, L, B, and A, while the lowest quantities were obtained from area F. The highest O contribution was found in sampling areas G, H, L, and C. The highest OS contribution was recovered from areas C, B, E, A, and F. The results of the statistical analysis performed for these sampling areas are presented in **Table 1**. In particular, OS was identified as a major contributor in two traces collected from area C. Lastly, 47% of the traces presented one or more unknown contributors.

Table 1. Statistical analysis of the traces recovered from the areas with greater OS contribution, presented in descending order. The following values are reported: percentage of traces suitable for comparison, percentage of traces that returned an LR for OS contribution lower than 10^6 and equal to or greater than 10^6, OS profile completeness, and OS relative contribution to the trace.

Sampling Area	Suitable	$1 < LR_{OS} < 10^6$	$LR_{OS} \geq 10^6$	OS profile completeness	OS mixture proportion
C	100%	28.6%	14.3%	25 - 81%	0.10 - 0.63
B	71.4%	28.6%	14.3%	63 - 88%	0.11 - 0.28
E	28.6%	28.6%	0%	66 - 69%	0.16 - 0.45
A	57.1%	14.3%	14.3%	47 - 75%	0.10 - 0.45
F	28.6%	14.3%	14.3%	50 - 69%	0.26 - 0.27

4. Discussion

The findings are in accordance with the simulated shooting scenario and serve to corroborate the case-work results. The greatest quantity of the gun owner's DNA can be retrieved from the barrel (areas G and H) and the magazine (area L), which is consistent with O loading the firearm prior to the shooting.

The OS can be successfully retrieved from the firearm's body (areas A, B, and C). The trigger/internal trigger guard (areas E and F) yield a lower DNA amount and less probative STR profiles. The Occasional Shooter rarely contributes more than the Owner to the trace: the Occasional Shooter was found to be the major contributor in only two traces, both of which were collected from the grip panels (area C) - the greatest area of hand-to-gun contact.

Despite the personal use of the gun, non-prevalent DNA was recovered from at least half of the traces. It is possible that the foreign DNA resulted from a secondary or higher-degree transfer, as evidenced by the identification of the partner of one participant as a contributor.

5. Conclusions

Provided that they are not the owner, the last individual to discharge a semi-automatic handgun can be successfully retrieved from the firearm's body. The grip

panels in particular exhibit the highest amount of deposited DNA. The last shooter was identified as a major contributor to the grip panels in 28.6% of the sampled firearms. Nonetheless, the owner is consistently the major contributor to the traces. Additionally, despite the firearm being exclusively owned and used by a single individual, non-prevalent DNA is frequently recovered. In conclusion, the characteristics of traces recovered from firearms appear to be predominantly determined by the deposition of DNA over time rather than by the last use of the item.

Institutional Review Board and Informed Consent Statement

This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the University of Perugia (protocol code 1051062, approved on 13 April 2022). Informed consent was obtained from all subjects involved in this study.

Funding

The authors did not receive external funding.

Declaration of competing interest

None.

References

[1] Polley, D.H., Mickiewicz, P., Vaughn, M.L., et al. An Investigation of DNA Recovery from Firearms and Cartridge Cases. Canadian Society of Forensic Science Journal, 39 (2006) 217 - 228. doi.org/10.1080/00085030.2006.10757145.

[2] Gosch A, Euteneuer J, Preuß-Wössner J, Courts C. DNA transfer to firearms in alternative realistic handling scenarios. Forensic Sci Int Genet. 48 (2020) 102355. doi.org/10.1016/j.fsigen.2020.102355.

[3] Kaesler T, Kirkbride KP, Linacre A. Improvements, factors, and influences on DNA recovery from firearms. Forensic Sci Int Genet. 65 (2023) 102873. doi.org/10.1016/j.fsigen.2023.102873.

[4] Pääbo S. Amplifying Ancient DNA, in PCR Protocols: A Guide to Methods and Applications. Academic Press; San Diego, CA, USA, 1990, pp.159-168.

[5] Genetisti Forensi Italiani Ge.F.I. Recommendations for Personal Identification Analysis by Forensic Laboratories (2018) Available online: 2004202091522.pdf (gefi-isfg.org).

[6] Bleka Ø, Storvik G, Gill P. EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. Forensic Sci Int Genet 21 (2016) 35-44. doi.org/10.1016/j.fsigen.2015.11.008.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 649-654
DOI: https://dx.doi.org/10.15304/cc.2025.1869

DNA on palmar and dorsal side of hands, and transfer of DNA during a slap and a punch of a face

Sean Galea[1], Annalisa Durdle[1,2], Roland A.H. van Oorschot[3,4*]

[1] *School of Life and Environmental Sciences, Deakin University, Geelong, Australia.*

[2] *Institute of Environmental Science and Research Limited, Auckland, New Zealand*

[3] *Office of Chief Forensic Scientist, Victoria Police Forensic Services Department, Macleod, Australia.*

[4] *School of Agriculture, Biomedicine and Environment, La Trobe University, Bundoora, Australia*

** Corresponding author at: Victoria Police Forensic Services Centre, 31 Forensic Drive, Macleod 3086, Australia. Email address: roland.vanoorschot@police.vic.gov.au*

Abstract

Little is known of the DNA quantities and relative proportions of self and non-self DNA on different areas of the dorsal side of hands compared to the palmar side. Further, there is little known regarding the bidirectional transfer when a hand grabs an arm, slaps a face with the palmar side of a hand, or punches a face with the dorsal side of a hand. Here we undertake two preliminary studies to start addressing these knowledge gaps. Study 1 compared the DNA quantities and profiles of samples collected from four areas on the palmar side and four areas of the dorsal side of the left and right hands of four participants. Study 2 considered the profiles collected from areas of an arm after being grabbed, areas of a face after being slapped by the palmar side of a hand and being punched by the dorsal side of a hand, as well as the extent and location of any DNA transferred to the hands during these actions. Study 1 showed that the quantity and proportion of self and non-self DNA vary among the different palmar and dorsal areas of hands. Study 2 showed that DNA can transfer bidirectionally between contacting areas of hands and another individual's body during grabbing, slapping and punching activities. More data is required to substantiate these initial findings and to generate probabilities of profile types generated from samples taken from areas of contact during grabbing, slapping and punching to assist investigations of alleged incidents of such activities.

Keywords

DNA transfer, hand, face, skin, punch, slap

Introduction

Touch DNA is often collected to assist investigations of alleged criminal activities, often from surfaces contacted by hands. While most hand contacts of interest relate to the palmar side of a hand, there are some where the dorsal side is of relevance, such as during investigation of a coward's punch, also known as a sucker punch, king hit or one-punch attack.

There is awareness of the quantities of DNA, the proportional presence of self versus non-self DNA, present on different areas of the palmar side of hands [1,2], and what is transferred from and to the palmar side of hands when contacting a surface, e.g.: [3-8]. However, there is a lack of this information in respect to the dorsal side. Further, while there are a few studies focused on the transfer, persistence and recovery of DNA from touched areas of skin other than the hand [9-13], there is also a general lack of information regarding skin-to-skin DNA transfer. We've conducted two preliminary studies to start filling this knowledge gap. *One study focused on the quantities and origins of DNA on different zones of the palmar and dorsal sides of hands (study 1) and another study focused of the bi-directional transfer between hands and grabbed arms as well as between hands and face after a slap with the palmer side of a hand, and after a punch with the dorsal side of a hand (study 2).*

Materials and methods

Four individuals participated in both studies. Sampling from zones 1 to 8 of hands (Figure 1), biceps of arms and the face was conducted using viscose swabs applying a wet/dry double swabbing method. In study 2, prior to the contact activities, samples were collected from areas adjacent to the areas of the arm and face to be contacted, i.e., anterior forearm and lower jaw (left and right side), respectively. After the sequential contact of the arms and face (Figure 3), samples were collected from the contacted areas of the arms and face as well as zones 1 to 8 of the contacting hands.

DNA was extracted from the swabs, quantified, amplified, profiled and deconvoluted as per standard casework procedures previously described [8]. An inclusion LR cut-off of 1000 was used, calculated without correcting for θ (co-ancestry co-efficient) or population substructure.

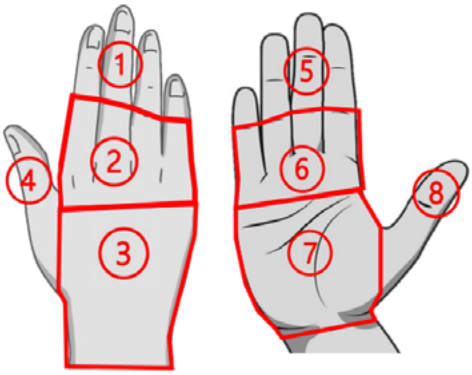


Figure 1. Zones 1-8 of a hand targeted for DNA sampling.

Results

Study 1

An investigation of the quantity of DNA, and composition of the profiles generated, from samples taken from Zones 1–8 of the left and right hand of four participants, showed that most DNA was collected from zones 1, 7 and 6; and the least DNA was collected from zones 2, 8 and 3 (Figure 2). Profiles were generated from 60 of 64 (94%) samples: the donor was sole or major contributor in 55 of 60 (92%) profiles; 48 of 60 (80%) profiles included an unknown; unknown individuals were the sole contributors in 5 samples – all on dorsal side (P1 4x, P3 1x).

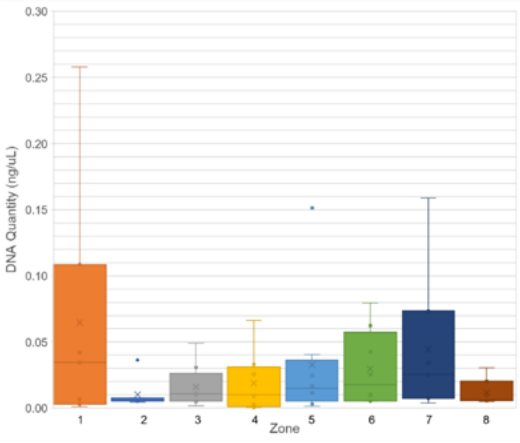


Figure 2. DNA quantities per zone of hand.

Study 2

An investigation of the detectability of direct, indirect, and bidirectional transfer of DNA during a mock assault (the grabbing of biceps followed by slapping or punching of a face) (Figure 3).

Pre-contact control samples were collected from areas adjacent to contacted areas of P2, P3, P4. Only self and unknown source DNA was detected in these samples.

Post-contact samples were collected from areas of contact from P2, P3, P4 and hand zones 1 to 8 of P1. Figure 3 shows the profile contributions detected in these samples. Direct DNA transfer was observed: from 'victim's' face to 'offender's' hands in 2 slap (zones 5+6) and 1 punch (zone 2) experiments; from 'offender' to 'victim's' face in 1 punch experiment; from 'offender' to 1st biceps 3 of 4 times; from 'offender' to 2nd biceps 1 of 4 times; from 2nd biceps to 'offender's' hands 3 of 4 times (zones 5,6,7,8). Indirect DNA transfer by the 'offender' was not observed.

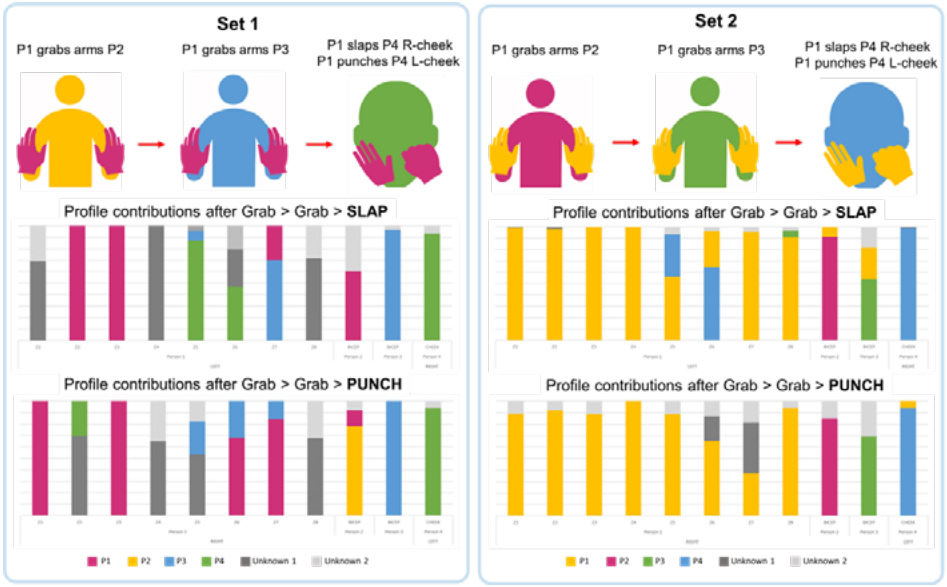


Figure 3. Sequence of contacts of hands grabbing biceps and slapping or punching a face and profile information of sample recovered from biceps, face and zones 1-8 of hands, for two test sets involving four individuals. Each of the four individuals (color coded) were in a different position within the sequence of contacts in set 1 and set 2.

Discussion

Study 1 shows that there is variation in the quantity and composition of DNA present in different areas of the palmar and dorsal sides of a hand.

Study 2 shows that detectable levels of DNA were transferred bidirectionally during a grab of an arm, a slap to a face and a punch to a face. These transfers were specific to particular parts of the hand. Further, no indirect transfer of P2 to P3 or P4, or of P3 to P4, by P1 was detected within this small study.

This information may assist investigators of particular alleged skin-to-skin assaults, in respect to sample targeting and activity level evaluations. These were preliminary studies with outcomes warranting additional studies of this kind to be conducted to gain further insights into the variation of DNA quantities and origins on different parts of the hand and the bidirectional transfer during their contact with areas of another individual's body. Further studies are also needed to generate probabilities of profile types generated from samples recovered from hands and body areas after particular types of skin to skin contacts.

Conclusion

The quantity and proportion of self and non-self DNA vary among the different palmar and dorsal areas of hands. DNA can transfer bidirectionally between contacting areas of hands and another individual's body during grabbing, slapping and punching activities. Awareness of this may offer opportunities to identify persons of interest by targeted sampling of particular areas of hands and bodies after particular types of alleged activities of interest. However, further studies are required to generate probabilities of profile types generated from samples taken from areas of contact during grabbing, slapping and punching to assist investigations of alleged incidents of grabbing, slapping and punching activities.

Acknowledgments

We thank the volunteers for their participation in this study.

Conflict of interest statement

No conflicts of interest.

References

[1] A.A. Oleiwi, M.R. Morris, W.M. Schmerer, et al., The relative DNA-shedding propensity of the palm and finger surfaces. Sci. Justice,55 (5) (2015) 329-334.

[2] D.L. McColl, M.L. Harvey, R.A.H. van Oorschot, DNA transfer by different parts of a hand. Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) e29-e31.

[3] A.K. Buckingham, M.L. Harvey, R.A.H. van Oorschot, The origin of unknown source DNA from touched objects. Forensic Sci. Int. Genet. 25 (2016) 26-33.

[4] B. Szkuta, K.N. Ballantyne, R.A.H. van Oorschot, Transfer and persistence of DNA on the hands and the influence of activities performed. Forensic Sci. Int. Genet. 28 (2017) 10-20.

[5] B. Szkuta, K.N. Ballantyne, B. Kokshoorn, et al., Transfer and persistence of non-self DNA on hands over time: using empirical data to evaluate DNA evidence given activity level propositions. Forensic Sci. Int. Genet. 33 (2018) 84-97.

[6] C.M. Cale, M.E. Earll, K.E. Latham, et al., Could secondary DNA transfer falsely place someone at the scene of a crime? J. Forensic Sci. 61 (1) (2016) 196-203.

[7] A. Cahill, L. Volgin, R.A.H. van Oorschot, et al., Where did it go? A study of DNA transfer in a social setting. Forensic Sci. Int. Genet. 73 (2024) 103101,

[8] M. Zacher, R.A.H. van Oorschot, O. Handt, et al., Transfer and persistence of intruder DNA within an office after reuse by owner. Forensic Sci. Int. Genet. 73 (2024) 103130.

[9] P. Wiegand, M. Kleiber, DNA typing of epithelial cells after strangulation. Int. J. Legal Med. 110 (4) (1997) 181-18.

[10] G.N. Rutty, An investigation into the transference and survivability of human DNA following simulated manual strangulation with consideration of the problem of third party contamination. Int. J. Legal Med. 116 (3) (2002) 170-173.

[11] K.G. de Bruin, S.M. Verheij, M. Veenhoven, et al., Comparison of stubbing and the double swab method for collecting offender epithelial material from a victim's skin. Forensic Sci. Int. Genet. 6 (2) (2012) 219-223

[12] Z.E. Bowman, K.S.A. Mosse, A.M. Sungaila, et al., Detection of offender DNA following skin-to-skin contact with a victim. Forensic Sci. Int. Genet. 37 (2018) 252-259.

[13] A.E. Fonneløp, H. Johannessen, G. Heen, et al., A retrospective study on the transfer, persistence and recovery of sperm and epithelial cells in samples collected in sexual assault casework, Forensic Sci. Int. Genet. 43 (2019) 102153,

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 655-663
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Dna transfer in the washing machine. Analysis with different detergents

Manuel Lozano-García[1*], Laura Jiménez Villegas[3], Sara Bravo Gómez[1], Sara Palomo-Díez[1,2], Cláudia Gomes[1,2], Eduardo Arroyo-Pardo[1,2], Ana María López Parra[1,2]

[1]*Legal Medicine, Psychiatry and Pathology Department, Medicine School, Complutense University of Madrid, 28040 Madrid, Spain.*

[2]*Forensic Sciences Group, Forensic Genetics and Toxicology, Health Research Institute of the Hospital Clínico San Carlos (IdISSC), 28040 Madrid, Spain*

[3]*Immunology, Ophthalmology and ENT Department, Medicine School, Complutense University of Madrid, 28040 Madrid, Spain.*

** Corresponding author at: Legal Medicine, Psychiatry and Pathology Department, Medicine School, Complutense University of Madrid, 28040 Madrid, Spain, manueloz@ucm.es,*

Abstract

Different studies have shown the existence of DNA transfers in washing machines. Transfer can be direct or through a vector, which will be indirect.

This study aims to analyse the effect of detergents with different pHs on direct and indirect transfer in washing machine washes. For this purpose, blood samples were taken intravenously in a tube without anticoagulants from two individuals, a man and a woman of similar ages. A 2x2 centimeters cotton cloth was added to 50µl of this blood, which was placed in the washing machine with a clean cloth with the same characteristics. All the fabrics had previously been irradiated for 48 hours with UV and then were impregnated with blood on the same day. The washing time has been considered variable, so T0 corresponded to the first wash and T6 to the sixth washing day. After washing the fabrics were left to dry for 48 hours.

A total of 80 washes were performed, 40 with Ariel detergent (pH 7.4) and 40 with Kiriko (pH 8-9), using for both a fast cycle (1 hour) at 60ºC and with a total of 160 tissues, two for each wash (with blood and clean), for a total of 40 tissues per individual. Subsequently, DNA was extracted from the tissues by the FTA BioScience protocol (2022); (Whatman BioScience) and quantified by qPCR (Applied Biosystems™ Quantifiler™ Human DNA Quantification Kit). Finally, the data obtained were analysed with the SPSS v.29 program.

Significant differences in DNA concentrations were obtained between the detergents used (Ariel-Kiriko) and between stained and clean fabrics. The absence of quantification results with Ariel allows us to assume that there is some PCR inhibitory molecule among its components. It cannot be determined whether it also has a DNA-degrading effect. With Kiriko detergent, we obtained high DNA concentrations in both fabrics, although much higher in the stained fabrics.

Keywords
Transfers, detergents, blood, washing machines, fabrics, pH.

Introduction

When we put the clothes in the washing machine, you might think that it will eliminate any trace of dirt and matter that may have remained in the fabric of the garment. But, as it has been seen in different studies, [1,2,3,4] this could not be the case. In our case, we investigate blood stains that may appear on fabrics due to a crime. This blood can be detected after washing with chemicals such as luminol [1] or DNA from the blood can remain, making it possible to obtain a genetic profile [5]. Furthermore, this genetic material can pass between different clothes in the washing machine due to transfer [6].

DNA transfer is a process by which an individual's genetic material appears on an object or other person, through a possible vector, usually a third person. The issue arises from the fact the donor of the genetic material never had direct contact with the object in question (direct transfer). Instead, an intermediary, or vector, was involved, resulting in an indirect transfer [7,8,9,10].

These transfers can cause a DNA sample to appear at a crime scene and in our research, we will focus on the existence of washing machine DNA transfers, work that has already been demonstrated in different studies [2,3,4].

Specifically, our objective is to analyze the effect of detergents with different pH on the direct and indirect transfer in washing machines.

Material studied, methods, techniques

Samples

The blood samples were obtained from a man and a woman. Blood was extracted without additives from a vein, which was used simultaneously to stain the fabrics with 50 microliters. Blood sampling was done twice, on two Wednesdays in a row. The first week to stain the 40 fabrics that were washed with a specific detergent,

Ariel® (Procter & Gamble), and the following Wednesday blood was drawn again to stain the 40 fabrics that were washed with another detergent Kiriko® (Blue Liquid Detergent).

Variables

The variables considered in this study, which were compared with each other using the quantification data obtained, were:

— Type of fabric: We differentiate between clean and blood-stained fabrics in this group.
— Type of detergent: Compare the data of the fabrics that were washed with the Ariel® detergent, which had a pH of 7.4, versus those washed with the Kiriko® detergent, which had a pH of 8-9.
— Time that the fabric was stained until it was washed: The fabrics were stained on a Wednesday, if they were washed that same day, they were named T0, if they were washed on Thursday, they were named T1 (after the stained fabric had passed a day) and so on. the rest of the days; Friday = T2, Monday = T5 and Tuesday = T6.

Material and Methods

This research was carried out using 160 cotton fabrics. These were divided into two groups of 80, some would be used as clean cloths, and the others would be stained with 50 microliters of blood. From each group of 80, two groups of 40 would be made again, some would be washed with Ariel® detergent and others with Kiriko® detergent. Finally, of the group of 40 stained fabrics (both those washed with Ariel® and Kiriko®), 20 were a woman's blood and the other 20 were a man's blood. This explanation can be seen more schematically in Figure 1.

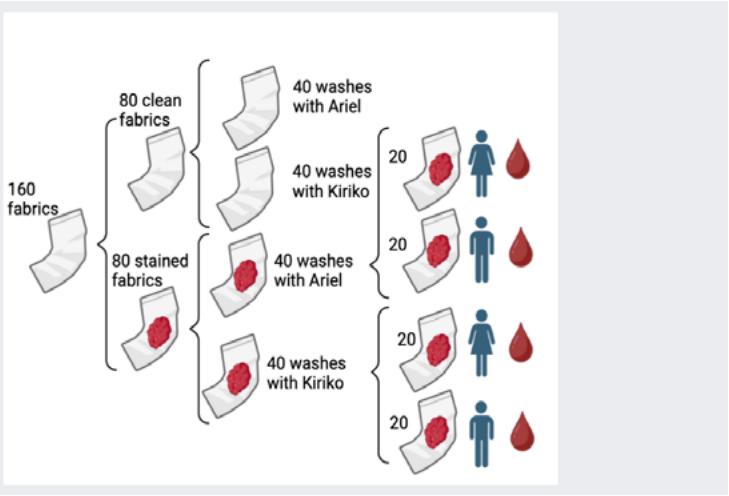


Figure 1. Scheme of the organization of fabrics used in the study. Image created by the author

A total of 80 washes were carried out. These were organized into 20 Ariel® washes with the man's blood samples, 20 Ariel washes with the woman's blood sample, 20 Kiriko washes with the man's blood sample, and another 20 Kiriko washes with the woman's blood sample. Each group of 20 with different detergents and blood sample were divided into 5 groups depending on the time the fabrics were stained with blood (T0, T1, T2, T2, T5, and T6). In this way, 4 replicates were obtained for each of the times.

The procedure that was carried out began with cutting out the 160 2X2 centimeter cotton fabrics. Before starting the research, these were exposed to 48 hours of ultraviolet light in a laminar flow hood. In this way, we ensured that any remaining exogenous DNA was eliminated.

After 48 hours, the fabrics that were stained with blood were selected as explained above and 50 microliters of blood were poured onto the fabrics. These clothes would be recognized by having a cut in one corner if it was the woman's blood and cuts in two corners if it was the man's blood. No corners were cut from the clean fabrics.

Next, they started with the washes in which a clean fabric was added along with a stained fabric in the washing machine. A rapid one-hour wash cycle at 60ºC was used for all washes. Between washes, it was interspersed with the same cycle with bleach. The detergents used were:

— Ariel PODS+ in capsules which contained: > 30% anionic surfactant, 5-15% soap, <5% non-ionic, phosphonates, enzymes, optical brightener, perfumes, benzyl salicylate, citral, hexyl cinnamal, limonene, linalool.
— Kiriko basic, linen which contained: <5% anionic surfactants, nonionic surfactants, amphoteric surfactants, soap, polycarboxylates, contains optical white enzymes, perfume (HEXYL CINAMAL), and Methylisothiazolinone, methylchloroisothiazolinone.

Once the washing was completed, the fabrics were left to dry for 24 hours, and the DNA was extracted from the fabrics. For this, the FTA BioScience protocol (2022) was used; (Whatman BioScience). A cutout from the center of the fabric was used for this protocol.

Finally, with the DNA extractions completed, the quantification of all samples was carried out using qPCR (Applied Biosystems™ Quantifiler™ Human DNA Quantification Kit).

The statistical analysis carried out involved calculating significance using the ANOVA test. The statistical analyses were conducted using IBM SPSS Statistics 22 software. The significance level was set at $p < 0.05$.

Results

Significant results have been obtained by comparing DNA quantifications between stained fabrics and clean fabrics with a value of $p<0.001$ and between fabrics washed with Ariel detergent versus those washed with Kiriko detergent with a value of $p<0.001$. In Tables 1 and 2 you can see the quantification data and the differences between detergents and fabrics.

No significant differences were found in the time variables, obtaining a value of $p<0.775$ in the comparison between the quantifications of the fabrics that spent more time with the blood versus those that spent less time.

Table 1. Quantification of clean fabrics washed with Kiriko and Ariel detergents. Legend: L1,2... refers to the order of washing. PL refers to clean fabric, I1 to individual 1 (man), and I2 to individual 2 (woman). Finally, T0 and T6 refer to the number of days that the fabrics were stained with blood.

Name	Kiriko® qPCR (ng/µL)	Ariel® qPCR (ng/µL)
L1 PL T0	Undetermined	Undetermined
L1 PL T6	0,000797	Undetermined
L2 PL T0	Undetermined	Undetermined
L2 PL T6	0,001733	Undetermined
L4 PL T0	Undetermined	Undetermined
L4 PL T6	0,000902	Undetermined

Table 2. Quantification of stain fabrics washed with Kiriko and Ariel detergents. Legend: L1,2... refers to the order of washing. PL refers to clean fabric, I1 to individual 1, and I2 to individual 2. Finally, T0 and T6 refer to the number of days that the fabrics were stained with blood.

Name	Kiriko qPCR (ng/µL)	Ariel qPCR (ng/µL)
L1 I1 T0	0,470526	Undetermined
L1 I1 T6	1,329528	0,000795
L2 I2 T0	0,649014	Undetermined
L2 I2 T6	0,258205	Undetermined
L4 I2 T0	0,176519	Undetermined
L4 I2 T6	0,236353	0,001465

Discussion

As can be seen in the results (Tables 1 and 2), washing a fabric does not eliminate the blood although not visually. Furthermore, as expected, there is a greater amount of DNA in fabrics that are stained by blood than in clean fabrics, although it has been proven that there is DNA transfer due to the presence of genetic material on clean fabrics

Regarding the types of detergents, according to the data it could be said that Ariel® detergent is much more effective when washing fabrics than Kiriko® detergent because the quantifications are much lower in fabrics washed with Ariel®. However, observing an agarose gel with genetic material from samples washed with Ariel® (figure 2), a different color of the bands can be seen. These samples look

intense blue, compared to the positive control that is found bottom right which has a greener color. This made us think that perhaps Ariel® was not so good at degrading and removing DNA, but rather had some type of PCR inhibitor. For this reason, it cannot be determined if it has a degrading effect on samples with DNA.

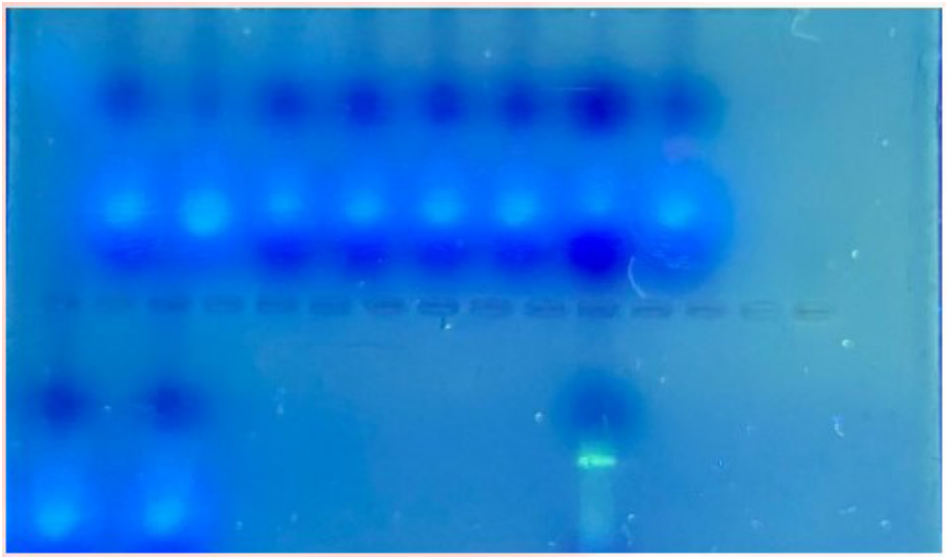


Figure 2. Agarose gel from samples washed with Ariel®. Blue fluorescence is seen in the samples and green is the positive control. This gel shows us the possibility of PCR inhibition in the samples.

Conclusion

In conclusion, the cases of genetic material transfer between fabrics in the washing cycles can be confirmed although a much more representative sample is needed, considering that some individuals "donate" more material genetic than others, due to dandruff, flaking, and skin problems, among others. On the other hand, the influence of the detergents' pH on the DNA degradation cannot be affirmed, since further research is needed on the possible inhibition in the PCR in the cases of fabrics washed with Ariel® detergent.

Acknowledgments

Thank the Department of Legal Medicine, Psychiatry and Pathology of the Faculty of Medicine of the Complutense University of Madrid and especially the Forensic Genetics and Population Genetics laboratory for making this research possible.

I also want to thank Laura Jimenez Villegas for the great work she did in the laboratory during her internship and that led us to have these results.

Conflict of interest statement

There is no conflict of interest.

References

[1] Edler, C., Gehl, A., Kohwagner, J., Walther, M., Krebs, O., Augustin, C., Klein, A., (2017). Blood Trace Evidence on Washed Textiles - a systematic approach. Int J Legal Med. 131(4): 1179 1189.

[2] Kamphausen, T., Fandel, S. B., Gutmann, J. S., Bajanowski, T., & Poetsc, M. (2015). Everything clean? Transfer of DNA traces between textiles in the washtub. Journal of Forensic Sciences, 129:709–714 DOI 10.1007/s00414-015-1203-5.

[3] Roland van Oorschot, A. H., Szkuta B., Meakin G.E., Kokshoorn B., Goray M., (2018). DNA transfer in forensic Sciencie: A review Forensic Science International: Genetics

[4] Roland van Oorschot, A. H., Meakin, G. E., Kokshoorn, B., Goray, M., & Szkuta, B. (2021). DNA Transfer in Forensic Science: Recent Progress towards Meeting Challenges. Genes 12, 1766. https://doi.org/10.3390/genes12111766

[5] M. van den Berge, G. Ozcanhan, S. Zijlstra, A. Lindenbergh, T. Sijen, (2016). Prevalence of human cell material: DNA and RNA profiling of public and private objects and after activity scenarios, Forensic Sci. Int. Genet. 21: 81–89.

[6] Lowe, A., Murray, C., Whitaker, J., Tully, G., Gill, P. (2002). La propensión de los individuos a depositar ADN y la transferencia secundaria de ADN de bajo nivel de individuos a superficies inertes, Forensic Sci. Ent. 129 (1): 25–34, https://doi.org/10.1016/S0379-0738(02)00207-4.

[7] Thornbury, D., Goray, M., Van Oorschot R. (2021). Transfer of DNA without contact from user clothing, pillowcases and towels by shaking agitation. Science and Justice. 61: 797-805.

[8] M. Goray, E. Eken, RJ Mitchell, RA van Oorschot, (2010). Transferencia secundaria de ADN de sustancias biológicas en diferentes condiciones de prueba, Forensic Sci. En t. Gineta, 4 (2): 62–67, https://doi.org/10.1016/j.fsigen.2009.05.001.

[9] Van Oorschot, R.A.H., Szkuta, B., Meakin G., Kokshoorn, B. (2019). DNA transfer in forensic science: A review. Forensic Science International: Genetics. 38: 140-166.

[10] Verdon, TJ., Mitchell, RJ., Van Oorschot, RA,. (2013). La influencia del sustrato en la transferencia de ADN y la eficiencia de extracción, Forensic Sci. En t. Gineta, 7 (1): 167–175, https://doi.org/10.1016/ j.fsigen.2012.09.004.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 665-672
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Investigation into the Effects of Gunshot Residue on STR PCR Inhibition in DNA Testing from Firearms

Christian G. Westring[1, 2 *], Elizabeth Kopitke[3], Toni M. Diegoli[4], Todd Bille[4]

[1] *Niagara County Sheriff's Office, Forensic Laboratory, Lockport, NY - USA*

[2] *Purdue University Northwest, Department of Biology, Hammond, IN - USA*

[3] *Independent Forensics, Hillside, IL – USA*

[4] *United States Bureau of Alcohol, Tobacco, Firearms, and Explosives, National Laboratory Center, Beltsville, MD – USA*

** Corresponding author at: Niagara County Sheriff's Office, Forensic Laboratory, 5526 Niagara Street, Lockport, NY 14094 – USA. Tel: +1 (716) 4383313 Email: christian.westring@niagaracounty.com*

Abstract

In 2021, firearms were implicated in 85.7% of homicide cases, highlighting the critical need for law enforcement to enhance forensic investigative capabilities regarding touch DNA evidence recovered from firearms. Forensic DNA laboratories routinely gather touch DNA from various parts of firearms, such as grips, slide serrations, magazines, triggers, and other components, to facilitate the identification of individuals involved in criminal activities. However, the impact of inhibitory compounds from gunshot residue (GSR) on DNA analysis using modern technologies remains inadequately understood. Firearms-related samples often yield complex DNA profiles, posing challenges for forensic analysts. These profiles frequently manifest as mixtures involving three or more individuals, yet little is known about the potential effects of co-eluting metallic compounds from DNA extraction on polymerase chain reaction (PCR) efficiency and/or inhibition levels. This lack of understanding extends to the physical location on the firearm (grips vs. slide serrations, etc.) and the presence of GSR during DNA collection, which may vary in dosage and inhibit PCR reactions to varying degrees.

Our current study addresses these gaps by examining the varying levels of PCR enhancement and/or inhibition associated with touch DNA collections from firearms. We aim to illustrate how the quantity of GSR-related inhibition varies based on the site of sample collection and the analytical methodology employed for DNA profiling. While advancements in DNA typing technologies have improved sensitivity, these systems are still susceptible to limitations imposed by inhibitory substances co-eluting with DNA during extraction and purification. By identifying

and quantifying metallic species from the primer, cartridge cases, bullet construction, and other potential GSR combustion products present in samples collected from firearms, forensic practitioners can refine DNA sampling techniques and biochemical testing methods. This approach aims to enhance the quality and reliability of touch DNA profiles recovered from firearms, thereby aiding law enforcement agencies in criminal investigations.

Keywords
DNA; Firearms; Forensic Science; Touch DNA, PCR

1. Introduction

In 2021, firearms were involved in 85.7% of homicide cases, underscoring the urgent need for law enforcement to strengthen forensic investigative capabilities, particularly concerning the recovery and analysis of touch DNA evidence from firearms. Recent advancements in the sensitivity of modern DNA technologies have significantly enhanced the feasibility of collecting trace DNA from firearms and ammunition, garnering considerable interest among forensic scientists [1-3]. Numerous studies have been conducted to evaluate optimal sample collection sites, DNA quantity and yield, as well as more effective methods for DNA collection [4-8]. However, the impact of DNA degradation and the inhibitory effects of gunshot residue (GSR) on DNA analysis using current technologies remains insufficiently explored [9-11]. Forensic analysts frequently encounter complex DNA profiles from firearms-related samples, often presenting as mixtures involving three or more individuals [12-16]. Despite this, the potential influence of co-eluting metallic compounds from DNA extraction on polymerase chain reaction (PCR) efficiency and inhibition levels remains poorly understood and warrants further investigation [17].

2. Material studied, methods, and techniques

2.1 Firearms and ammunition sample preparation

Gunshot residue (GSR) samples were generated using two distinct firearms frequently encountered in law enforcement applications both domestically and internationally: the Glock 17 (9mm) and the Heckler & Koch USP9 (9mm). Prior to firing, both firearms underwent a thorough cleaning process using a commercial solvent, Hoppe's No. 9 solvent in order to minimize traces of contamination. Following the cleaning, each firearm was used to discharge 50 rounds of Winchester 9mm 115-grain full metal jacket ammunition. Background GSR samples were subsequently collected using the conventional double-swab method with Puritan spun

cotton swabs (wooden shaft). Samples were taken from two specific locations on the firearms: 1) the front of the slide, from the area forward of the ejection port extending to the muzzle, and 2) the interior of the barrel. The collection process utilized a 2% sodium dodecyl sulfate (SDS) solution to aid in the efficient retrieval of GSR and trace DNA.

2.2 DNA Analysis – Extraction and Quantification

Swabs containing GSR collected from both firearms were carefully divided into equal halves to facilitate duplicate analysis (data not shown). For the initial analysis, the PrepFiler BTA™ Forensic DNA Extraction Kit (Thermo Fisher Scientific, Waltham, MA) was utilized to extract DNA from the samples. Following the extraction process, the DNA quantity was measured using the QuantiFiler Trio kit (Thermo Fisher Scientific, Waltham, MA) on the 7500 Real-time PCR System (Thermo Fisher Scientific). Before proceeding to amplification, all quantified samples underwent normalization to ensure consistent DNA content across the analyses.

2.3 DNA Analysis – STR Amplification, Detection, and Analysis

Data sets were generated using high-quality DNA, such as 63 pg of DNA from buccal swabs, combined with varying volumes of "GSR extract." In these experiments, between 1 and 14 µL of GSR extract was used in place of water to adjust the PCR reactions to their full volume. Each batch of samples was analyzed alongside a comprehensive set of controls, including positive controls, negative controls, and extraction blanks, to ensure the reliability of the data. Amplification of the DNA samples was performed using the GlobalFiler PCR Amplification Kit (Thermo Fisher Scientific, Waltham, MA). The amplified products were then analyzed through capillary electrophoresis on the Applied Biosystems 3500 Genetic Analyzer at the specified target injection times. STR profiles were visualized and interpreted using GeneMapper ID-X and OSIRIS genotyping software prior to further downstream analyses.

Results

To assess the impact of DNA extracts containing trace amounts of gunshot residue (GSR) on downstream PCR applications, four sets of swabs were processed using routine DNA STR analysis, similar to standard casework samples. Contrary to expectations, DNA extracts containing GSR, even when unquantified, did not exhibit any volumetric enhancement or inhibitory effects on PCR, as previously observed

in the author's laboratory. A more detailed locus-specific analysis (Figure 2) did not reveal any consistent cases of locus-specific inhibition, regardless of the firearm used or the volume of extract amplified. Furthermore, these effects were not more pronounced in low template samples (63 pg) compared to the stochastic amplification typically observed in samples containing less than 100 pg of DNA (Figure 3). This suggests that any amplification variations may align more with the inherent challenges of amplifying low template DNA rather than the presence of GSR in the extracts.

Discussion

Interestingly, the inhibitory effects previously observed with increasing volumes of GSR extract in other routine methods and chemistries were not detected in this study. Earlier results had shown an initial "boost" in PCR amplification when small volumes of GSR extract were added, compared to the control group, eventually leading to complete PCR inhibition with larger volumes of GSR (data not shown). The initial increase in amplification efficiency may indicate the presence of higher concentrations of metallic species such as Mg2+ [18-19], which is known to enhance PCR efficiency in small amounts. However, under higher concentrations of Mg2+, complete PCR inhibition has also been documented. The absence of any reproducible inhibitory effects from varying GSR extract volumes in this study raises several intriguing questions.

Do the new generation DNA extraction technologies more effectively remove PCR inhibitors that previously co-eluted with DNA? Alternatively, are the newer STR amplification kits better equipped to manage the potential inhibitory effects of GSR? Some insight into this issue is provided by studies conducted by Czado et al. (2022), which reported low levels of metallic contaminants, such as copper and zinc, following DNA extraction. Despite these findings, further research is required to fully understand the influence of GSR and metallic species—whether originating from GSR itself or from metallic cartridge casings—on downstream PCR applications. These questions remain crucial for advancing forensic DNA analysis in cases involving firearms.

Conclusion

Firearms are frequently submitted to forensic crime laboratories for DNA analysis, a process that has become highly standardized. However, a variety of collection

techniques, DNA extraction methods, and STR chemistries are routinely employed across different labs. In this study, we aimed to investigate the potential inhibitory effects of gunshot residue (GSR) on STR analysis. Previous research using DNA extraction techniques such as phenol-chloroform-isoamyl alcohol (PCIA) and various paramagnetic DNA isolation methods demonstrated a volume-dependent increase in PCR inhibition or enhancement due to GSR, which, in some cases, resulted in complete PCR inhibition.

In contrast, the current study did not observe the same patterns of PCR enhancement or inhibition with low template (63 pg) samples using modern STR chemistries. On both firearms, additional alleles beyond those matching the operator were detected. Notably, these firearms were new and had not been previously handled or fired by anyone other than the operator prior to cleaning and use for this study. Interestingly, DNA was successfully recovered from the interior of the barrel, matching the DNA profile of the individual who loaded the magazine before firing. This finding suggests that DNA can be transferred to internal firearm components during loading, potentially influencing forensic DNA analysis in firearms-related cases.

Acknowledgments

The authors wish to extend their sincere gratitude to Dr. Phillip Danielson at the University of Denver for his invaluable insights, discussions, and contributions to this project. Additionally, we would like to express our appreciation to Dr. Karl Reich of Independent Forensics for his continued support and contributions. We also thank Purdue University Northwest for providing the funding necessary to complete this project. It is important to note that the study sponsors had no role in the design of the study, the collection, analysis, and interpretation of the data, the writing of the manuscript, or the decision to submit the manuscript for publication.

Conflict of interest statement

The authors have no financial or personal interests with other people or organizations that could inappropriately influence (bias) the research conducted in this project.

References

[1] Goray, M., Mitchell, R.J., van Oorschot, R.A.H., Investigation of secondary DNA transfer of skin cells under controlled test conditions, Legal Medicine, 12:117-120, 2010

[2] Gosch, A., Euteneuer, J., Woessner, J.P, Courts, C., Analysis of DNA transfer to firearms considering relevant alternative handling scenarios, FSI Genetics Supplement Series 7:433-435, 2019

[3] Gosch, A., Euteneuer, Woessner, J.P, Courts, C., DNA transfer to firearms in alternative realistic handling scenarios, FSI Genetics, 48;102355, 2020

[4] Courts, C., Madea, B., Schyma, C., Persistence of biological traces in gun barrels – an approach to an experimental model, Int J. Legal Med, 126:391-397, 2012

[5] Kaesler, T., Kirkbride, K. P., Linacre, A., Improvements, factors, and influences of DNA recovery from firearms, FSI Genetics, 65:102873, 2023

[6] Kaesler, T., Kirkbride, K. P., Linacre, A., Persistence of touch DNA on commonly encountered substrates in different storage conditions, Forensic Science International, 348:111728, 2023

[7] Polley D., Mickiewicz, P., Vaughn, M., Miller, T., et. al., An investigation of DNA recovery from firearms and cartridge cases, Canadian Society of Forensic Science Journal, 2013

[8] Prasad, E., Hitchcock, C., Raymond, J., et. al., DNA recovery from unfired and fired cartridge cases: A comparison of swabbing, tape lifting, vacuum filtration, and direct PCR, Forensic Science International, 317: 110507, 2020

[9] Montpetit, S., O'Donnell, P., An optimized procedure for obtaining DNA from Fired and Unfired Ammunition, FSI Genetics, 17: 70-74, 2015

[10] Chang K.H, Jayaprakash, P.T., Yew, C.H., Abdullah, A.F.L, Gunshot residue analysis and its evidential values: a review, Australian Journal of Forensic Sciences, 45:1, 3-23, 2013

[11] Bill T.W., Fahrig, G., Weitz, S.M., Peiffer, G.A., An improved process for the collection and DNA analysis of fired cartridge cases, FSI Genetics 46, 2020

[12] Fan, G.Y., Li, Wei., Li, S.T., Zhang, Q.Y., An evaluation of the performance of DNA recovery from fire firearms and cartridge cases using microdialysis filtration, FSI Genetics Supplement Series, 2017

[13] Gerard, R.V., McVicar, M.J., Lindsay, E., Randall, E.D., Harvey, E., The long range deposition of gunshot residue and the mechanisms of its transportation, Canadian Society of Forensic Science Journal, 44:3, 97-104, 2011

[14] Fairley, M., Hall, D., A single approach to the recovery of DNA and firearms discharge residue evidence, Science and Justice, 44:1, 15-19, 2004

[15] Holland, M., M., Bonds, R. M., Holland, C. A., McElhoe, J. A., Recovery of mtDNA from unfired metallic ammunition components with an assessment of sequence profile quality and DNA damage through MPS analysis, FSI Genetics, 39;86-96, 2019

[16] Martin, B., Blackie, R., Taylor, D., Linacre, A., DNA profiles generated from a range of touched sample types, FSI Genetics, 36:13-19, 2018

[17] Martin, B., Plummer, A., Linacre, A., Henry, J., Direct PCR of fired shotgun casings: a South Australian evaluation, Australian Journal of Forensic Sciences, 2020

[18] Grunewald, H., Optimization of Polymerase Chain Reactions, Methods in Molecular Biology, 226: 89-99

[19] Ely, J. J., Daniel-Reeves, A., Campbell, M., Kohler, S., Stone, W. H., Influence of Magnesium Ion Concentration and PCR Amplification Conditions on Cross-Species PCR, BioTechniques, 25:1, 38-42, 1998

Sample Name	Description	Small Auto. (ng/µL)	Large Auto. (ng/µL)	Degradation Index	Small Auto. Ct	Large Auto. Ct	IPC Ct	DaQ
CW1		0.013	0.005	5.5	67.3	66.5	55.6	-
CW2		0.004	0.001	14.8	70.9	72.3	55.3	-
CW3		0.012	0.008	3.3	67.6	65.1	55.4	-
CW4		0.001	0.000	3.4	73.7	37.6	55.4	-
CW5		23.552	22.236	1.1	22.4	20.2	29.6	-
-	-	-	-	-	-	-	-	-

Figure 1. DNA quantitation results from swabs collected from CW1 – forward half of slide of Glock 17, CW2 – inside barrel of Glock 17, CW3 – forward half of H&K USP9, CW4 - inside barrel of H&K USP9, and CW – 5, reference buccal swab from test subject. DNA concentrations on both firearms were comparable across test sites (e.g. front of G17 and USP9 pistol, 13pg and 12pg respectively) after rigorous cleaning using commercial firearm solution (Hopes#9). NOTE: In both sets of barrel swabs, detectable quantities of human DNA were observed after firing.

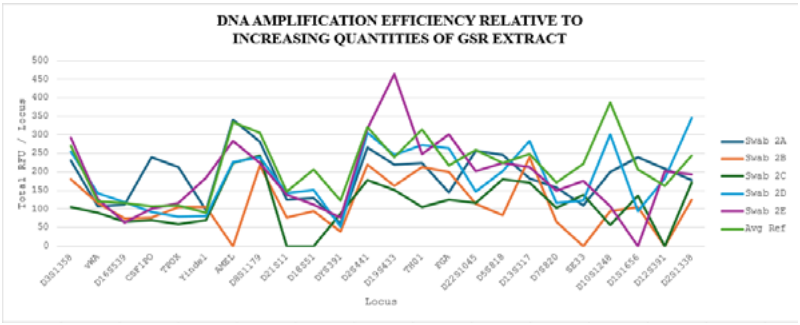


Figure 2. DNA extracts containing GSR were tested at varying input quantities (by volume) to determine the potential impact of GSR on PCR amplification inhibition for a) locus specific effects, and b) GSR quantity specific effects. Substantive difference were neither see for locus specific and GSR inhibition were noted in DNA samples ranging from low concentrations of GSR extract (1ul) to high concentrations (14ul) for samples amplified with 63pg of input DNA. For comparison, 63pg of "Reference" type DNA were amplified in triplicate (see Avg Ref data) at 63pg as well.

Figure 3. Amplification of DNA samples extracted by phenol chloroform isoamyl alcohol employing a previous generation STR kit (e.g. Promega PowerPlex16HS) demonstrated substantive volume dependent inhibition with greater and greater volumes of DNA extract when compared to comparable samples sans GSR. While the exact mechanism behind this inhibition remains elusive, it was not observed in the current study when employing a paramagnetic DNA isolation technique (PrepFiler) and the GlobalFiler STR chemistry. Future studies will further explore the reproducibility of these findings.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 673-679
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Development, characterization and qualification of a cellular target for tracking biological trace transfers in forensic applications

Stéphane Sauvagère[3], Mathilde Bastat[1,2], Sarah Bayol[1,2], Mathilde Recipon[1], Franck Carreiras[1], Olivier Gallet[1], Rémy Agniel[1], Cédric Picot[1], Joséphine Gibert[2], Sylvain Hubac[2], Johanne Leroy-Dudal[1], Sabrina Kellouche[1], Christian Siatka[3,4], Francis Hermitte[1,2*]

[1]*Equipe de Recherche sur les Relations Matrice Extracellulaire-Cellules, ERRMECe, Groupe MEC uP, CY Cergy Paris Université, Neuville sur Oise, France.*

[2]*Institut de Recherche Criminelle de la Gendarmerie Nationale, Cergy-Pontoise, France.*

[3]*École de l'ADN de Nîmes, Nîmes, France.*

[4]*UPR CHROME, Université de Nîmes, Nîmes, France.*

**Corresponding author: francis.hermitte@gendarmerie.interieur.gouv.fr*

Abstract

During forensic investigations, "Touch DNA", invisible traces transferred through physical contact, is a challenge. At the Institut de Recherche Criminelle de la Gendarmernie Nationale in France, Touch DNA represented 80% collected samples but 50% genetic profiles were unexploitable. A better understanding of the dynamics of their transfer is crucial to optimize their management. We have explored the potential of cell-derived lipids as markers of residual cell-fragments susceptible to contain DNA, to track "touch DNA" traces. Lipid dyes enabled the detection of residual cell-fragments from in vitro models of human keratinocytes and calibrated fingermarks using microscopy or Crime-lite®Auto. This method allows for monitoring cell-derived fragments transfer during swabbing, opening the way to an adaptable technique for both traces' detection and transfer tracking. The labelling strategy did not interfere with genetic analysis; it could be a new application in forensic techniques.

Keywords

Touch DNA, fingermarks, keratinocytes, lipids markers, biological traces transfer

1. Introduction

Forensic sciences have always had a crucial role in criminal investigations. Biological traces were essential for investigation due to their potential for DNA profiling[1]. However, to obtain genetic profiles from specific biological traces, known as "Touch DNA," is a real challenge for investigators[2]. These traces, transferred through physical contact, may be invisible and /or contain limited amounts of biological material [3]. A better understanding of transfer dynamic mechanisms may enable an improvement detection of touch DNA and thus may increase yield of exploitable genetic profiles. Recent studies have shown that cell-derived markers persist for over 3 months on fragmented and/or dead corneocytes, the main type cells found in contact deposition [4]. This suggests that they could serve as effective targets for tracking deposits [5]. Lipids are a potential target because they are found especially in the epidermis as an integral component of cellular membranes. They constitute a reservoir of markers surrounding corneocytes [6]. The application of lipid markers in forensic science offers several advantages such as lower toxicity compared to DNA markers; and they can be detected using non-destructive techniques, such as fluorescence microscopy [7].

This work proposes to use Nile Blue, Nile Red lipid dyes, known for their applications in fingerprints detection, to visualize smaller deposits in both *in vitro* and fingermarks models [8]. By combining this approach with traditional DNA analysis, we aim to contribute to a more comprehensive understanding of "touch DNA" transfer mechanisms.

2. Materials and methods

2.1. Models of "touch DNA"

For in vitro models, 25,000 keratinocytes (KH) from human skin cell line (CRL-2309™, ATCC) were seeded on coverslip (Epredia™, Thermo Fisher Scientific) in serum-free keratinocyte medium (Thermo Fisher Scientific) containing L-glutamine, supplemented with recombinant human epidermal growth factor and bovine pituitary extract. After 4 hours allowing cell adhesion, the medium was removed and the cells were dried at room temperature (RT) during 1 hour. For fingermarks, we develop a device, allowing volunteers (all samples were obtained with the informed consent of the volunteers and the approval of the ethics committee of the Pôle Judiciaire de la Gendarmerie Nationale (PJGN), in

accordance with the Helsinki Accords (1975) and the French National Charter of Research Integrity) to apply their finger on a microscope glass (Epredia™, Thermo Fisher Scientific) with a calibrated pressure, for 10 seconds. Samples were then left to dry 30 minutes at RT. In some experiments, samples were swabbed 15 seconds after the swab was moistened (milliQ water). Swabs were stored at -20°C prior to analysis.

2.2. Lipid labelling

Nile Red (NR, Invitrogen) was prepared at 500 µg/mL in acetone. Nile Blue (NB, Thermo Fisher Scientific) was prepared at 1 mg/mL in milliQ water. For experiments, NR and NB solutions were diluted respectively in PBS. Then, traces models were incubated with 500 µL of solution during 10 or 20 minutes for NR and NB respectively. Samples were washed three times with 500 µL of PBS and counterstained with Hoechst (33342, Thermo Fisher Scientific) at 1:3000. Samples were mounted with Prolong Gold Antifade (P36930, Thermo Fisher Scientific) and imaged using a confocal laser scanning microscope (LSM900, Zeiss). Moreover, the Crime-lite® Auto (Foster + Freeman Ltd) was used in a darkroom.

2.3. Compatibility of labelling with genetic analyses

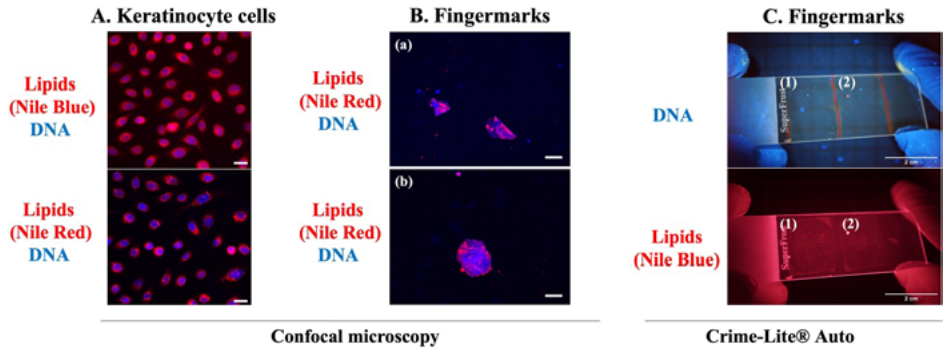
Cells were collected using 4N6FLOQ genetic swab (4504C, Copan). DNA was extracted using the Crime Prep Adem-kit®, Automag solution (Ademtech). The lysate was purified using KingFisher Duo Prime® (Thermo Fisher Scientific). DNA genotyping was carried out using the GlobalFiler IQC PCR Amplification kit (Thermo Fisher Scientific). 15 µL of samples was added to the multiplex PCR reaction mix. Positive control for PCR was DNA at 0.1 ng/µL and negative control was extra pure water. Amplification was performed in an Applied Biosystems™ Veriti™ Thermal Cycler, 96-well (Thermo Fisher Scientific) according to manufacturer's program. The multiplex PCR targets 21 autosomal STRs, a Y chromosome STR, a polymorphic insertion/deletion marker on the Y chromosome, the amelogenin gene and includes an Internal Quality Control (IQC).

For genotyping, samples were analyzed using ABI 3500 XL Genetic Analyzer for Human Identification sequencer (Applied Biosystem). Obtained results were analyzed using GeneMapper ID-X v1.6 software. Peaks with a fluorescence lower than 50 rfu, minor "off-ladder" (OL) peaks, stutters lower than 150 rfu and lower than 15% of the main peak and "drop-ins" lower than 400 rfu were removed from the profile. The number of peaks for each genetic profile was recorded and compared to that of the controls. The quality sensors ratio was calculated. If the value

obtained exceeded 1, the sample was considered degraded and/or inhibited. If the value was lower, the sample was considered unaltered. For each allele independently, the average of the replicates was calculated. The Welch T test was applied to compare the total average of labeled samples versus that of the controls. The confidence interval was set at 95%.

3. Results

3.1. Validation of lipid markers

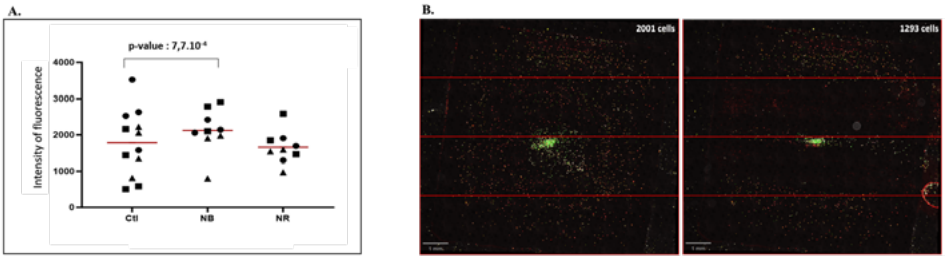


(a) and (b) represented two different fingermarks. Scale bar: 20 µm. (1) and (2) represented one fingermark made by middle finger successively. Scale bar: 2cm.

Figure 1. Visualization of lipids labelling on in vitro keratinocyte cells model (A) or calibrated fingermarks (B) with confocal microscopy. Observation of lipids labelling on calibrated fingermarks with Crime-lite® Auto (C).

Using microscopy, lipids labelling efficiently enabled the detection of in vitro keratinocyte cells model (Fig. 1A). For fingermarks, corneocytes were well labeled with lipid dyes. This labelling was heterogeneous and punctiform (Fig. 1B). Using Crime-lite®Auto (Fig. 1C), the Hoechst alone (DNA marker) did not allow for the distinction of cell-derived elements within fingermarks. However, lipids staining has revealed punctuate elements in fingermarks apposition. Thus, lipids are good candidate for tracking cell-derived fragments within deposits, especially Nile Blue which can highlight cell-fragments using Crime-lite®.

3.2. Interests of lipids staining for tracking traces



For (2A): Unlabeled (Ctl) or Nile Blue (NB) or Nile Red (NR) labelled keratinocyte cells were collected with the 4N6FLOQ Swab®. The average fluorescence for each profile was reported and analyzed with a Welch's t-test. The black squares, triangles and circles represent the experimental replicates and the red line represents the average of the experiments. For (2B): Labelled cells on a glass slide were swabbed and remaining cells were counted. Scale bar: 1cm. Red lines represented different collected areas.

Figure 2. Study of the influence of lipid markers on genetic analysis (A) and the benefits of lipid staining for counting cells after collect (B).

The impact of the lipid markers on the PCR was assessed by calculating the quality sensors ratios of each profile. Out of the 60 profiles, none had a ratio higher than 1 indicating neither inhibition of the PCR (data not show). The average fluorescence of the profiles obtained with or without lipid probes (NB, NR) fells within a range of emission suitable with genetic profiling. Additionally, there was no decrease of fluorescence with probes compared to the control, indicating no interference with genetic analysis (Fig. 2A). A slight increase of the fluorescence emission of the genetic profiles was observed with NB, and should be considered in further experiments (Fig. 2A). Finally, applying lipid staining to deposits of cells on supports could allow monitoring their sampling and evaluating the number of collected cell fragments (Fig. 2B).

4. Discussion

This work has presented a strategy for biological traces investigation using lipids markers revealing and tracking cell-derived fragments within contact deposition. This method has offered the advantage of fast (less than 1h), cost-effective, and less toxic process compared to direct DNA detection which could be toxic. Moreover, it allows for targeting cell fragments that are larger than DNA molecule, and thus

improving the chances of collect biological traces. Finally, it could serve as a valuable non-invasive and less destructive method that could preserve the integrity of the biological trace for subsequent DNA analysis [7]. In the future, the influence of the physicochemical properties of surfaces where traces are deposited should be explored, as these could affect detection [9]. Furthermore, the persistence of lipids stains on cell fragments has to be studied to use them as tracker to follow traces transfer.

5. Conclusion

Lipid labelling could be used to highlight elements of cellular origin in the biological traces and to track them in order to better understand their dynamics. Finally, this staining is of real interest to users in the field, as it is completely safe.

6. Acknowledgments

We would like to thank the French National Agency for Scientific Research (ANR) for financially supporting this research project as part of the BIOTRACK project (ANR-21-CE39-0020). We would also like to thank the EDSI Doctoral School at the CY Cergy Paris Université for actively supporting this project. Confocal microscopy experiments were carried out at 'Microscopies & Analyses' imaging core, I-Mat, Cergy Paris Université. The authors would like to thank the Conseil régional Île-de-France for funding the COMCOI@CY project (grant: EX061044) that has allowed for the acquisition of a ZEISS LSM900. Finally, we would like to thank all the partners in the BIOTRACK consortium: Dr. Vincent CASTELLA, Séverine NOZOWNIK, Michel CHAPRON, Petr DOBIAS, Cordula HAAS, Mohamed-Amine KHELIF, André PILLIN, Etienne PILLIN, Olivier ROMAIN, for their support during this work.

7. Conflict of interest

The authors declare no conflict of interest.

8. References

[1] Sijen T, Harbison S. On the Identification of Body Fluids and Tissues: A Crucial Link in the Investigation and Solution of Crime. Genes (2021); 12:1728.

[1] Tozzo P, Mazzobel E, Marcante B et al. Touch DNA Sampling Methods: Efficacy Evaluation and Systematic Review. Int J Mol Sci (2022); 23:15541.

[1] Quinones I, Daniel B. Cell free DNA as a component of forensic evidence recovered from touched surfaces. Forensic Sci Int Genet (2012); 6:26–30.

[1] Burrill J, Daniel B, Frascione N. A review of trace "Touch DNA" deposits: Variability factors and an exploration of cellular composition. Forensic Sci Int Genet (2019); 39:8–18.

[1] He M, Zhou X, Wang X. Glycosylation: mechanisms, biological functions and clinical implications. Sig Transduct Target Ther (2024); 9:1–33.

[1] Bouwstra JA, Nădăban A, Bras W et al. The skin barrier: an extraordinary interface with an exceptional lipid organization. Prog Lipid Res (2023); 92:101252.

[1] Recipon M, Agniel R, Leroy-Dudal J et al. Targeting cell-derived markers to improve the detection of invisible biological traces for the purpose of genetic-based criminal identification. Sci Rep (2023); 13:18105.

[1] Prabakaran E, Pillay K. Nanomaterials for latent fingerprint detection: a review. Journal of Materials Research and Technology (2021); 12:1856–85.

[1] Recipon M, Agniel R, Kunemann P. et al. Detection of invisible biological traces in relation to the physicochemical properties of substrates surfaces in forensic casework. Sci Rep (2024); 14:13271.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 681-688
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Interpretation and education: (sub) source and activity issues, similar challenges?

Tacha Hicks[1*], Alex Biedermann[2], Christophe Champod[2], Franco Taroni[2]

[1] *Fondation pour la Formation Continue Universitaire Lausannoise, Lausanne, Switzerland.*

[2] *University of Lausanne, Faculty of Law, Criminal Justice and Public Administration, School of Criminal Justice, Switzerland.*

** Corresponding author at: School of Criminal Justice, Batochime, 1015 Dorigny, Switzerland, tacha.hickschampod@unil.ch, tel. +41 21 692 46 07*

Abstract

In this paper we present what we believe to be the mechanisms for success in the acquisition skills and competence in evaluation and reporting. We discuss the different means of online and certifying education we developed to tailor the curriculum to the role of the learner. At one extreme, there are Massive Open Online Courses (MOOC) or short single-topic video clips to raise awareness among scientists and the public. At the other end of the spectrum are certifying, in-depth, longer learning courses with individual feedback and tutoring. The latter type of course prepares DNA scientists for the challenges of forensic interpretation in casework, including reporting and testimony, and research.

As practical examples of problems best addressed by the latter type of education, we present theoretical and practical concepts that can help scientists formulate more meaningful propositions. We also discuss problematic examples of reporting that are often observed in the context of alleged activities. More generally, we illustrate the similarities in the challenges DNA scientists face in reporting the value of their findings, regardless of the hierarchical level of the propositions.

Although much progress has been made in recent years with the use of specialised software (particularly probabilistic genotyping), publications and court transcripts indicate that the concept of likelihood ratio used in and produced by such software is still misunderstood by many forensic scientists. This illustrates that education in forensic DNA interpretation by academics specialising in the field is both necessary and timely. We argue that the continuing education of DNA scientists/researchers and the training of key players (investigators, prosecutors, defence lawyers, judges) through the creation of flexible learning pathways should be a central part of tomorrow's forensic science landscape.

Keywords
Evaluative reporting, principles of interpretation, e-learning, MOOC, communication.

Introduction

It is now widely accepted that forensic scientists should reason in the face of uncertainty using a robust framework of logic. Much effort has been devoted to the development of research [1-11] and guidelines/recommendations [12-15] to promote an approach to interpretation that is based on a scientific measure of uncertainty, considers the views of both parties, sustains scrutiny and results in transparent reporting. However, the guidelines are difficult to implement because they require in-depth knowledge and understanding.

In this paper we present our work on different ways to acquire skills and competences in evaluation and reporting that we believe can help to implement current recommendations and guidelines. Our learning opportunities are based on flexible pathways that allow participants to learn at their own pace, depending on their personal and professional constraints. In addition, to meet the needs of the community, different types of courses have been developed so that the curriculum can be tailored according to one's role. At one end of the spectrum, we offer a Massive Open Online Course (MOOC) to raise awareness of the importance of DNA evidence in court. At the other end of the spectrum, we offer in-depth, longer learning courses with individual feedback and tutoring. The latter type of course can prepare DNA scientists for the challenges of forensic interpretation in casework, including reporting and testimony, and research.

A first example of online training: MOOC "Challenging forensic science; how science should speak to court"

The aim of our MOOC "Challenging forensic science; how science should speak to court"[1] is to promote critical thinking about forensic science. It is designed to alert (without alarming) investigators, prosecutors, defence lawyers, judges, scientists and the public (e.g., journalists) to the limitations of forensic science methods and techniques to promote the sound administration of forensic science in the criminal justice system. Through videos about famous cases (causes célèbres), the course emphasises the importance of probabilistic reasoning in forensic science. This 16-hour MOOC is freely available on the Coursera platform

[1] https://www.coursera.org/learn/challenging-forensic-science

and is offered in both in English and French, with subtitles in 21 languages. Since its launch in January 2019, more than 17'000 individuals from all over the world have enrolled.

This MOOC is divided into five modules (e.g., weeks). In the first week, entitled "What is the "DNA" of a good forensic report?", learners explore the criteria that an inferential framework should meet when reasoning in the face of uncertainty. The module highlights the differences between evaluative statements and other types of forensic reports (e.g., technical reports), while introducing the principles of forensic evaluation [e.g., 1-5] and how to assign the value (i.e., LR) of forensic observations and findings. The second week, "Elementary: source is not activity!", covers the distinction between assessing the value of forensic results (e.g., DNA, gunshot residues) when the issue is not only the source of the material but also the activities from which it potentially resulted. Week three, "DNA is not the magic bullet", focuses on DNA-related cases where the meaning of the results has been misunderstood. In week four, "Statistics in Court", participants study cases such as Dreyfus, Clark and Collins, where statistics were misapplied. The course concludes with week five, "The Wonderland of Certainty", which examines cases such as Dallagher (earmarks), Mayfield and McKie (fingermarks). It also discusses what is at stake when an expert decides to conclude to an identification/individualisation (i.e., source attribution determination).

The MOOC is built around high-profile cases and includes interviews with scientists, lawyers and individuals directly involved in these cases.

University-based certifying online courses for reporting forensic scientists

Competence in forensic evaluation and reporting is difficult to acquire. Attending workshops or MOOCs is useful to alert forensic DNA scientists to what to avoid and how to improve. However, exercises and individual tutorials are needed to fully master concepts such as probabilities, likelihood ratios or the formulation of propositions. Over the past 15 years, the University of Lausanne has developed online courses on the evaluation of DNA results given propositions at different hierarchical levels (i.e., at sub-source level and activity level). Since 2009, some 200 students from all continents (Africa, Asia, Europe, North and South America and Oceania) have enrolled. In these courses, participants practise formulating propositions, learn how to evaluate results in complex situations (also with probabilistic graphical models), and how to report their conclusions. Some examples of interpretation and reporting challenges are discussed below.

Why training matters: examples of recurrent interpretation challenges.

a. Formulating propositions

In our experience, DNA forensic scientists often find it difficult to distinguish between propositions and explanations, although there have been many publications on that topic [3,9,11-13]. The key point is that propositions are based on case information and must be formulated so that DNA results can help the court resolve a key issue (e.g., whether Mr Smith is the source of the DNA or if it is an unknown person). In contrast, explanations are based on the results and are appropriate for exploring potential reasons for specific observations (e.g., investigating who could be the source of the DNA). When evaluating YSTR DNA profiles, a common problem is the formulation of propositions such as "The male DNA is from Mr Smith or someone from his paternal line" versus "The male DNA is from an unrelated person". It would be fine to rely on the DNA analysis results and methods to explain the results and provide investigative leads. But if the investigation shows that there is no alternative source of the DNA in the paternal line, then there is no case information to justify this proposition. Also, if one alternate source is a paternal relative, then this should appear in the alternative proposition, not in the main proposition. Propositions are not determined by the method: they are dictated by the (key) issue(s) which in turn direct(s) the choice of method(s). The fact that individuals in the same paternal line can be expected to have certain similarities in their YDNA profile is part of general knowledge. It is worth disclosing and explaining to close (or not) investigation avenues. If the investigation shows that a relative is a viable alternative, then new methods ideally need to be devised (e.g., RM-YSTR). If it is not possible, then one should take into consideration this possibility in the alternative proposition. But, this should not interfere with the formulation of the proposition considering that Mr Smith is the source of the male DNA. Indeed, it is not the paternal line that is on trial, but Mr Smith.

Formulating propositions, when there is disagreement about how or when the DNA was transferred, is also difficult. In such a situation, again, it is often tempting to "explain" the findings *ex post* (i.e., *after* knowing the typing results) by saying, for example, that the DNA was (in)directly deposited by a particular mechanism. However, with such a formulation, it is difficult to evaluate the DNA results, and the probability of the results will be one, if they are perfectly "explained". Moreover, as outlined in [16], if the event "transfer" or "DNA deposition" is woven into the proposition, because the factfinder assesses the probability of propositions (usually informally), this means that the factfinder is left with the task of accounting for the phenomenon of DNA transfer (i.e., they will need to assess the probability of DNA being (in)directly deposited by a particular mechanism). We do not

think that laypersons have the knowledge to manage this complex issue. Scientific aspects such as transfer, prevalence, or background, are factors that are part of the evaluation of the results and as such should *not* be part of the proposition. It follows that a proposition must not explicitly state that DNA was deposited (e.g., "The person of interest discarded the knife *and that's how his/her DNA got on the knife*"). A more meaningful proposition is one that includes *only* the alleged activities, for example "stabbing" and not the alleged transfer mechanisms. This also has the advantage of allowing to assess all results, absence of DNA included.

b. Joint evaluation of autosomal and non-autosomal results

It is sometimes thought that DNA forensic scientists should not perform a joint evaluation of autosomal and non-autosomal results. However, it is questionable whether this task should be left to the fact finder. It has been shown that laypersons do not know how to combine probabilities or likelihood ratios obtained from different evaluative reports. Combining different types of forensic biological findings is not trivial and requires special knowledge and skills. Therefore, this task should be the domain of forensic DNA scientists. A meaningful couple of propositions that would allow examiners to jointly evaluate autosomal and non-autosomal DNA results could be: "Mr Smith is the source of the DNA", "An unknown unrelated person is the source of the DNA". As previously outlined, the alternative source (an unrelated person or not) would be based on the available case information.

c. What a likelihood ratio is and what it is not

Publications and court transcripts indicate that the concept of likelihood ratio is still misunderstood by many forensic scientists. It is not uncommon to read confusing statements such as "the LR is basically a division; the likelihood ratio equals proposition one divided by proposition two".[2] However, although it is a ratio, a likelihood ratio (or Bayes Factor) is, in the simplest case, obtained by dividing the probability of the results given one proposition by the probability of the results given the alternative. It is the probability of the findings or results that scientists are concerned with, not the propositions nor their probabilities [6].

At present, many scientists, speak of "sub-source LR" or "activity LR". While this might be seen as a convenient shortcut in informal conversations, it does not help communication. Indeed, if we want to communicate that we can give a pro-

[2] All examples of problematic statements in this paper are paraphrased from actual examples encountered by the authors in their teaching, casework and research.

fessional opinion on our results, but not on the propositions themselves, then it is confusing, if not misleading, to use language that suggests that the scientist is describing the value of the *propositions* (e.g., "*activity* LR"), rather than the results or findings.

Asserting that a likelihood ratio "supports one proposition more than the other" is another example of misrepresenting the value of the evidence. While it is true that LRs are numbers, it is the *results*, not the LRs, that support one proposition over the other (assuming the LR is different from one).

Another common error, even among experienced forensic scientists, is to say, for example: "The probability of observing this DNA profile is at least a billion times *more likely* if the DNA mixture is from Mr Smith and three unknown, unrelated individuals than if it is from four unknown, unrelated individuals." The problem with this sentence is that it qualifies a probability as "likely", which is confusing. It amounts to placing a probability on a probability. One way to avoid this confusion is to say: "The DNA comparison result is of the order of a billion times more probable if the DNA mixture is from Mr Smith and three unknown, unrelated individuals than if it is from four unknown, unrelated individuals."

Conclusions

In this paper we have reviewed various educational frameworks and highlighted challenges in interpretation and reporting that we believe reflect a lack of specialised education on this topic.

With our MOOC, we aim to make the recipient of forensic information aware that uncertainty should be managed by using probabilities and applying the principles of interpretation. The "Essentials of DNA Interpretation" course aims to provide DNA forensic scientists with an appropriate theoretical and practical background in probabilistic and statistical reasoning so that they can tackle challenging DNA casework (e.g., formulation of propositions, comparisons of a DNA profile mixture to multiple persons of interest, assess complex DNA profiles) and report the value of their results in a robust manner, taking into account recent publications in DNA interpretation. The course "Advanced DNA Interpretation" enables participants to acquire specialised and up-to-date knowledge in the evaluation of forensic biological results when accounting for transfer, persistence, prevalence and recovery (TPPR) of biological trace material in the context of alleged activities.

Conflict of interest statement

Tacha Hicks is employed part-time by the *Fondation pour la Formation Continue UNIL-EPFL* (Foundation for Continuing Education of the University of Lausanne and the Swiss Federal Institute of Technology, https://www.formation-continue-unil-epfl.ch), which offers several online courses in forensic interpretation.

References

[1] I.W. Evett, B.S. Weir, Interpreting DNA Evidence—Statistical Genetics for Forensic Scientists, Sinauer Associates, Inc., Sunderland, MA, USA, 1998.

[2] C. Aitken, F. Taroni, S. Bozza. Statistics and the Evaluation of Evidence for Forensic Scientists, third ed., John Wiley & Sons, Chichester, UK, 2021.

[3] G. Jackson, The Development of Case Assessment and Interpretation (CAI) in Forensic Science. Ph.D. Thesis, University of Abertay, Dundee, UK, 2011.

[4] J.S. Buckleton, J.-A. Bright, D. Taylor, Forensic DNA Evidence Interpretation, second ed., CRC Press, Boca Raton, FL, USA, 2016.

[5] B. Robertson, G.A. Vignaux, C.E.H. Berger, Interpreting Evidence, second ed., Wiley & Sons, Chichester, UK, 2016.

[6] I.W. Evett, Avoiding the transposed conditional, Sci. Justice. 35 (1995) 127–131.

[7] R. Cook, I.W. Evett, G. Jackson, et al., A hierarchy of propositions: Deciding which level to address in casework. Sci. Justice. 38 (1998) 231–240.

[8] I.W. Evett, G. Jackson, J.A. Lambert et al., The impact of the principles of evidence interpretation on the structure and content of statements. Sci. Justice. 40 (2000) 233–239.

[9] I.W. Evett, G. Jackson, J.A. Lambert, More on the hierarchy of propositions: Exploring the distinction between explanations and propositions. Sci. Justice. 40 (2000) 3–10.

[10] G. Jackson, S. Jones, G. Booth et al., The nature of forensic science opinion—A possible framework to guide thinking and practice in investigations and in court proceedings. Sci. Justice. 46 (2006) 33–44.

[11] T. Hicks, J. Buckleton, Vincent Castella, et al., A logical framework for forensic DNA interpretation. Genes, 13 (2022) 957. https://doi.org/10.3390/genes13060957.

[12] S.M. Willis, L. McKenna, S.D. McDermott, et al., ENFSI Guideline for Evaluative Reporting in Forensic Science: Strengthening the Evaluation of Forensic Results

Across Europe (STEOFRAE). 2015. http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf, 2015 (accessed 6th of October 2024).

[13] P. Gill, T. Hicks, J.M. Butler et al., DNA commission of the International society for forensic genetics: Assessing the value of forensic biological evidence—Guidelines highlighting the importance of propositions: Part I: Evaluation of DNA profiling comparisons given (sub-) source propositions. Forensic Sci. Int. Genet. 36 (2018) 189–202.

[14] P. Gill, T. Hicks, J.M. Butler et al., DNA commission of the International society for forensic genetics: Assessing the value of forensic biological evidence—Guidelines highlighting the importance of propositions. Part II: Evaluation of biological traces considering activity level propositions. Forensic Sci. Int. Genet. 44 (2020) 102186.

[15] Expert Working Group on Human Factors in Forensic DNA Interpretation (2024). Forensic DNA Interpretation and Human Factors: Improving Practice Through a Systems Approach. National Institute of Standards and Technology, Gaithersburg, MD, NIST IR 8503. https://doi.org/10.6028/NIST.IR.8503

[16] T. Hicks, A. Biedermann, de Koeijer JA, et al., The importance of distinguishing information from evidence/observations when formulating propositions. Sci. Justice. 55(2015) 520-5.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 689-695
DOI: https://dx.doi.org/10.15304/cc.2025.1869

The ENFSI ReAct project: results of a large consortium investigation to standardise methods related to activity issues

Peter Gill[1,2*], Ane Elida Fonneløp[1,3], Tacha Hicks[4,5], Stavroulla Xenophontos[6], Marios Cariolou[6], Roland van Oorschot[7,8], Iris Buckel[9], Ingo Bastisch[9,10] and the ReAct Consortium

[1]*Department of Forensic Sciences, Oslo University Hospital, Oslo, Norway*

[2]*Department of Forensic Medicine, University of Oslo, Oslo, Norway.*

[3]*Centre for Ecological and Evolutionary Synthesis (CEES), University of Oslo, Norway.*

[4]*School of Criminal Justice & Fondation pour la formation continue universitaire lausannoise, University of Lausanne, Switzerland.*

[5]*Forensic Genetics Unit,University Center of Legal Medicine, Lausanne—Geneva, Lausanne University Hospital and University of Lausanne, Switzerland.*

[6]*The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus.*

[7]*Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, Macleod, Australia.*

[8]*School of Agriculture, Biomedicine and Environment, La Trobe University, Bundoora, Australia.*

[9]*Bundeskriminalamt, Wiesbaden, Germany.*

[10]*International Commission on Missing Persons, The Hague, The Netherlands.*

** Corresponding author at: Forensic Genetics Research Group, Oslo University Hospital, Oslo, Norway, peterd.gill@gmail.com*

Abstract

The ReAct (Recovery, Activity) project is an ENFSI (European Network of Forensic Science Institutes) supported initiative comprising a large consortium of laboratories. Here, the results from more than 23 laboratories are presented. The primary purpose was to design experiments simulating typical casework circumstances; collect data and to implement Bayesian networks to calculate likelihood ratios given activity level propositions. Two different experimental designs were used to simulate a robbery, where a screwdriver was used to force a door or window. The experiments were designed following laboratory feedback listing typical casework circumstances.

Keywords
ReAct, Activity issues, transfer experiments, ENFSI, Bayesian network.

Introduction

Over recent years, there has been much more awareness about reporting given activity issues, and court-attending scientists often give evidence to help address this. Scientists will resort to the body of literature to make generalisations, and there are many examples of this - for reviews see [1, 2].

However, there has been very limited work to investigate reproducibility of tests between laboratories, although [3, 4] examined this across 2-4 laboratories.

The difficulty is compounded by lack of data available for independent scrutiny; the receiver of the information must rely upon isolated peer reviewed material.

However, published papers primarily speak to experiences of a particular laboratory. This does not mean that these same experiences are applicable to different laboratories, especially those that run different systems and processes.

The ReAct project, was instigated as a collaborative exercise to generate a body of data from at least 23 laboratories with the following objectives:

a. To identify two kinds of cases, typically encountered in casework and agree on a common experimental protocol to be conducted by all participating labs.
b. To generate an open-access database of results.
c. To produce open-source software that can form the basis of a repository for a collection of Bayesian Networks.
d. To examine reproducibility between different laboratories.
e. To gain experience with coordinating large projects with multiple participant laboratories.
f. To describe the associated challenges.

Material studied, methods, techniques

Experiments undertaken

Data were collected for cases that simulated a burglary using a screwdriver. Three separate experiments were undertaken (*sampling to profiling carried out according to per lab routine protocol*):

a. The first was a control experiment, where a screwdriver was grasped, and the laboratory carried out tests to determine recovery of DNA from the handler, along with unknown contributors.
b. An individual (A) used screwdriver for 5 minutes as the first handler. It was then used by a second individual (B) as the second handler (direct transfer) for 2 minutes. DNA recovery of first, second and unknown contributors was studied (*note that same timings were used in subsequent experiments*). It should be noted that the disputed activities are very similar.
c. Individual A shook hands with individual B, individual B used a screwdriver, and the DNA recovery of A (handshaker), along with contributor B (last handler) and unknown contributor DNA was calculated. These experiments involved recovery of DNA from contributor B given an activity involving possible direct transfer, from contributor A given possible secondary transfer, and from unknown contributor(s) due to possible background (i.e., DNA present for unknown reasons). These experiments were repeated at three different time intervals relative to the shaking hands activity: 0, 1, 2 hours (experiments 3.0, 3.1 and 3.2, respectively). In the interim period, individual B resumed normal activities. There is a prior expectation that the contribution of contributor A will decrease over time.

Each experiment was replicated 20 or more times with different volunteers.

Software and databases

The results were collated and analysed with an R application, ShinyRFU(), that utilises EuroForMix [5], to assign mixture proportions (M_x) of contributors. Each laboratory determined quantity of DNA recovered; combined with M_x per contributor, the probability of DNA recovery for each contributor was calculated. These values were used directly in Bayesian networks to compute likelihood ratios. A package written in Shiny called ShinyReact() has been made available, along with all data collected from 23 laboratories. These data were used to inform Bayesian networks for the latter two experiments 2 and 3. Users of the software can choose a dataset to analyse, define the input Bayesian network parameters and calculate likelihood ratios for their findings, along with bootstrap derived confidence intervals. A user manual is available. Results are also compared to a qualitative method based on mixture proportions without DNA recovery attribution. A huge amount of data has been collected. In fact, a total of 2,700 DNA profiles have been analysed

and all these data are available for independent scrutiny. This work represents a massive undertaking without any parallel in the literature. The main purpose is to summarise the current work and to describe resources available (*see Resources section below*).

Results and Discussion

A large difference in the quantities of DNA recovered by the different laboratories, for all experiments, was observed (between 0.2 and 5 ng), depending on the sampling and recovery protocol.

Experiment 2

a. If an individual handles a screwdriver, then there is approximately 95% probability of DNA transfer, persistence and recovery (*abbreviated to recovery in subsequent text*). If two individuals handle a screwdriver, then the probability of recovery of the first handler drops to 88% (*mean values across all laboratories are quoted throughout*).
b. In approximately 43% of cases, the last handler aligned with the 'Major' (Mx>0.7), compared to 8% of cases for the first handler. Likelihood ratios were calculated where propositions addressed first or last handler. When a person aligns with the 'Major' contributor are low (LR≈1). However, observing a single DNA profile aligning with the person of interest (which happened in 8% of the ground truth experiments for the last handler, and less than 1% for the first handler), usually supports the view that the person of interest is the last handler rather than the first.

Experiment 3

a. As with experiments 1 and 2, there is approximately 95% probability of DNA recovery from the last handler.
b. DNA recovery from the handshaker in experiments involving indirect transfer was observed in c.69% of cases provided that the screwdriver was held immediately after handshake, otherwise it dropped to c.48% percent one hour after hand shaking: incidences were usually low level (only 4% aligned with 'Major' at time zero; dropping to less than 1% for one or two hours since handshake).
c. Incidences of results not supporting ground truth were restricted to single examples in experiments 3.1 and 3.2 where the former was a major

profile with low LR given activity level propositions. Because of the number of experiments undertaken and the magnitude of the LR, this is expected [6, 7]. Consequently, if the alternative proposition implies some form of social activity at least one hour previously between the person of interest and an alternative offender, a 'Major' (mixed) or 'Single' DNA profile aligning with the person of interest will usually support the proposition that the person is the handler, rather than an unknown person with whom he/she has socialised.

Resources

The main online resource is https://sites.google.com/view/altrap/enfsi-react-project. This website has links to all of the online resources which includes:

a. A summary of the ReAct project
b. Links to the ShinyReact() program and user manual
c. ShinyRFU() program access and user manual available at: https://sites.google.com/view/altrap/average-rfu-method
d. Links to excel spreadsheet databases:
e. Grand_compilation.xlsx of all data compiled into a single excel spreadsheet
f. Laboratory specific folders containing data and the complete analysis using the ShinyReact() program
g. All resources copied to GitHub: https://github.com/peterdgill/ShinyReact

An example of the output from ShinyReact() is shown in Table 1.

ng DNA recovery from the POI (known contributor)

z	0.01	0.5	1	1.5	2	2.5	3
0	20.4	106.2	207.8	323.2	451.4	591.6	743.2
0.05	0.4	2.2	4.3	6.7	9.3	12.3	15.4
0.1	0.3	1.7	3.4	5.2	7.3	9.6	12.1
0.15	0.3	1.4	2.7	4.2	5.9	7.8	9.7
2	0.0	0.1	0.2	0.3	0.4	0.5	0.6
3	0.0	0.0	0.1	0.1	0.2	0.3	0.3
5	0.0	0.0	0.0	0.1	0.1	0.1	0.1

Single contributor only

ng recovery from POI+Unknown contributor

Table 1. Tabulated likelihood ratios : Example from experiment 3, showing LRs for a single contributor model where recovery ranges from 0.01 – 3ng DNA, compared to recovery where a mixture of POI and unknown(s) are found. Note LRs are much greater where single contributors are recovered.

Conclusion

There was huge variation in DNA recovery across laboratories – some recovered c. 200pg, whereas others recovered c. 5ng. We can generalise that, in experiment 2 ground truth experiments, a 'Single' DNA profile aligning with the POI leads to LRs >1 when they are the only handler rather than the first and unknown person being the last handler. In experiment 3, LRs>1 are recovered if the person aligns with the 'Major' contributor when they are the handler, rather than having socialised with the true handler under the defence proposition. However, further work is needed if laboratory A wants to fully adopt the findings of laboratory B as the quantities between laboratories vary. To forward this aim, the focus of ReAct II is to study recovery efficiency, in addition to transfer and persistence, so that evaluation can incorporate the former as a laboratory dependent factor which can be incorporated into Bayesian Networks.

Acknowledgments

Funding for this Collaborative Project was received from the European Union's Internal Security Fund -Police (ISFP) – with Grant Agreement title: Competency, Education, Research, Testing, Accreditation, and Innovation in Forensic Science" [CERTAIN-FORS] ISFP-2020-AG-IBA-ENFSI and number: 101051099.

We acknowledge support from ENFSI and the ENFSI DNA Working Group Chairman Sander Kneppers

Conflict of interest statement

No conflicts of interest

References

[1] D. Taylor, B. Kokshoorn, A. Biedermann, Evaluation of forensic genetics findings given activity level propositions: A review, Forensic Sci Int Genet 36 (2018) 34-49.

[2] B. Szkuta, R. Ansell, L. Boiso, E. Connolly, A.D. Kloosterman, B. Kokshoorn, L.G. McKenna, K. Steensma, R.A.H. van Oorschot, DNA transfer to worn upper

garments during different activities and contacts: An inter-laboratory study, Forensic Sci Int Genet 46 (2020) 102268.

[3] K. Steensma, R. Ansell, L. Clarisse, E. Connolly, A.D. Kloosterman, L.G. McKenna, R.A.H. van Oorschot, B. Szkuta, B. Kokshoorn, An inter-laboratory comparison study on transfer, persistence and recovery of DNA from cable ties, Forensic Sci Int Genet 31 (2017) 95-104.

[4] B. Szkuta, R. Ansell, L. Boiso, E. Connolly, A.D. Kloosterman, B. Kokshoorn, L.G. McKenna, K. Steensma, R.A.H. van Oorschot, Assessment of the transfer, persistence, prevalence and recovery of DNA traces from clothing: An inter-laboratory study on worn upper garments, Forensic Sci Int Genet 42 (2019) 56-68.

[5] O. Bleka, G. Storvik, P. Gill, EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts, Forensic Sci Int Genet 21 (2016) 35-44.

[6] P. Gill, O. Bleka, T. Egeland, Does an English appeal court ruling increase the risks of miscarriages of justice when complex DNA profiles are searched against the national DNA database?, Forensic Sci Int Genet 13 (2014) 167-75.

[7] D. Taylor, J. Buckleton, I. Evett, Testing likelihood ratios produced from complex DNA profiles, Forensic Sci Int Genet 16 (2015) 165-71.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 697-704
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Evaluating the effects of environmental insults on presumptive and confirmatory tests and DNA recovery from blood stains

*Bushra Idris[1,2] and William Goodwin[1]

[1]*School of Law and Policing, University of Central Lancashire, Preston, UK*
[2]*Biotechnology Department, American University of Ras Al Khaimah, UAE*

** Corresponding author at: American University of Ras Al Khaimah, UAE, Busher.idris@aurak.ac.ae, 00971 528606070.*

Abstract

Blood deposit at crime scenes can be significantly impacted by local environmental factors such as UV radiation, temperature, and humidity. While it is well-established that environmental insults affect the efficacy of both presumptive and confirmatory tests there is limited published data on their empirical effects specifically related to blood samples.

A series of experiments were conducted using blood samples deposited on cotton, glass, and metal, which were then exposed to direct sunlight in Ras Al Khaimah (RAK), United Arab Emirates, over a 48-day period during spring. Samples were collected at intervals of 48 to 72 hours and subjected to two presumptive tests (Kastle-Mayer and Hemastix®) and two confirmatory tests (Hexagon® OBTI and RSID™-Blood). DNA quantity was assessed using RT-PCR and analysed with a 310 Genetic sequencer. The cumulative effects of temperature and time were measured using Accumulated Degree Days (ADD).

Presumptive tests consistently showed positive or weak positive results, while confirmatory tests began to fail, with the OBTI yielding negative results at 1215.5 ADD and the RSID at 693.5 ADD. DNA quantity decreased rapidly during the initial 475 ADD, followed by a gradual decline until 1425.5 ADD. Notably, blood samples on cloth and glass continued to produce full profiles in STR analysis using Identifiler® Plus. In contrast, blood samples on metal exhibited allele dropout starting at 943 ADD

Overall, comparing the results of presumptive and confirmatory tests with STR analysis revealed that presumptive tests can yield positive results even when DNA is insufficient for analysis. In contrast, confirmatory tests may return negative results despite the presence of amplifiable DNA but generally align more closely with downstream DNA analysis outcomes.

Keywords
Presumptive tests, confirmatory tests, bloodstains, STR analysis, DNA Degradation

Introduction

Blood is commonly encountered at crime scenes and correct identification is an integral part of the forensic analysis process. The presence of blood at a crime scene can significantly influence the outcome of a case, providing essential information for criminal investigations.

Environmental insults can hinder the identification of blood using both presumptive and confirmatory tests, as well as impact the quality and quantity of amplifiable DNA. Research has examined how factors such as temperature [1], humidity [2], soil composition [3], and ultraviolet radiation [4] influence DNA recovery from forensic samples. However, these investigations often occur in controlled laboratory settings, which may not accurately reflect the complex interactions found in natural environments.

While some studies have examined DNA persistence in biological samples [5], there is limited research on environments routinely exposed to extreme temperatures and UV radiation. Although it is acknowledged that environmental factors affect the effectiveness of presumptive and confirmatory tests, detailed empirical data on their specific impacts is limited.

We present a series of experiments investigating the effects of environmental insults on blood identification using both presumptive and confirmatory tests. Additionally, we compare these findings with the recoverable quantity of DNA in correlation to STR analysis.

Materials and methods

Experiment design

The experiment spanned 48 days outdoors, with sampling conducted every three days. A 50x50 cm white roof tile held three materials: glass microscope slides, 100% cotton cloth, and a stainless-steel kitchen knife. Each material featured marked 1 cm^2 grids, onto which 5 µl of blood was applied to five grids in each row, leaving a sixth grid empty as a negative control.

Prior to assembly, all materials were cleaned with 70% ethanol and rinsed with deionized water. For sampling, each 1 cm^2 grid area was either swabbed with

a wet cotton swab (for glass and metal) or cut out with a sterile scalpel for cloth. The collected samples were then placed in sterile 1.5 ml tubes and stored at -20 °C.

Accumulated Degree-Days (ADD)

Accumulated degree-days (ADD) were calculated from the daily hourly readings from RAK airport weather station 15 km away from the samples set up site [6]. ADD was calculated using the formula ADD = (Maximum + Minimum temperature)/2.

Presumptive and confirmatory testing

Blood samples were air-dried indoors for 2 hours before being moved to the experiment site. Samples from two grids for each material were collected immediately (Day 0) for positive controls. The Hemastix® and Kastle-Meyer (phenolphthalein) tests were performed as per manufacturer's instructions. Confirmatory tests RSID™-blood and Hexagon® OBTI were conducted following their procedures. Every 3 days, four samples from each material were collected and stored at -20 °C.

DNA extraction and quantification

Samples were collected from the 5th grid and DNA extraction was carried out using the Chelex-100 [7]. However, the incubation time at 56 °C was increased to 2 hours. Samples were then concentrated to 50 µl using microcon® filters (MilliPore®- Germany). Quantification was carried out using the Quantifiler™ Human DNA Quantification kit on an ABI 7500 real-time PCR machine, following the manufacturer's recommendations.

STR analysis

Samples from day 0, 17, 33, and 48 were selected to examine the effect of environmental insults on DNA profiles generated using the AmpF*l*STR® Identifiler® kit (Applied Biosystems) using a 310 Genetic Analyzer with GeneMapper™ ID version 3.2.

Results

Sample collection

Samples were collected every 3rd day from all material types. Some samples on metal and glass began to flake due to dehydration a few days into the experiment. In rows with such samples, the most intact grid was kept for DNA extraction.

Presumptive and confirmatory tests

The Hemastix® strips detected blood on all materials until Day 48, though results weakened over time. The phenolphthalein test showed similar trends, with weaker and slower results after Day 41.

In contrast, the OBTI test showed weak positives starting on Day 37 for metal, Day 41 for cloth, and Day 45 for glass. By Day 45, metal and cloth were negative, while glass had a very weak positive. The RSID test showed weak positives as early as Day 17 for metal and Day 29 for glass, and negative results by Day 29 for metal and Day 33 for glass. Blood samples on cotton remained positive until Day 41. All tests were positive up to Day 13 (Table 1).

Table 1: A table summarising the effect of environmental insults on the ability of presumptive and confirmatory tests on the identification of blood samples deposited on glass, metal and cloth.

DAY	ADD	HEMASTIX				p.phathaline				OBTI				RSID		
		G	M	C		G	M	C		G	M	C		*G*	*M*	*C*
13	368.5															
17	475															
21	582															
25	693.5															
29	809															
33	943															
37	1080															
41	1215.5															
45	1357															
48	1425.5															

*Dark shading (+v) Medium shading (weak +ve) Light shading (-ve)

*G= glass, M= metal and C= cloth.

DNA Quantification

A total of 42 samples were collected over the 48-day experiment. DNA content was compared for blood samples on metal, glass, and cotton in relation to ADD. Cotton had the highest DNA content, starting at 2.482 ng/µl on Day 0 and

decreasing to 0.252 ng/µl by Day 48. Glass samples started at 1.255 ng/µl and ended at 0.144 ng/µl, while metal samples showed the most significant drop, from 0.997 ng/µl to 0.054 ng/µl.

Metal samples experienced the fastest degradation, losing over 95% of the original DNA at an average rate of 2.1% per day. Glass samples followed with an 88.6% degradation rate, averaging 1.97% per day. Cotton retained DNA best, with only 82% depletion by Day 48, decreasing at a rate of 1.82% per day.

STR analysis

The STR profiles from blood samples on cloth showed a gradual decrease in DNA quantity, with a complete profile still visible at ADD 1425.5 (Day 48). The combined allele peak height (CPH) dropped from 41,916 at ADD 0 to 21,045 at ADD 1425.5.

Blood samples on glass also maintained a full profile at ADD 1425.5, though they showed slight peak imbalance at the FGA locus. The CPH decreased from 21,301 at ADD 0 to 14,576 at ADD 1425.5, lower than that of cloth samples.

In contrast, metal samples were most affected by environmental conditions. At ADD 1425.5, there was complete dropout of loci CSF and FGA, along with one allele dropout at loci vWA and D18. The CPH for metal samples plummeted from 43,694 at ADD 0 to just 4,358 at ADD 1425.5, indicating significant degradation compared to cloth and glass (Figure 1).

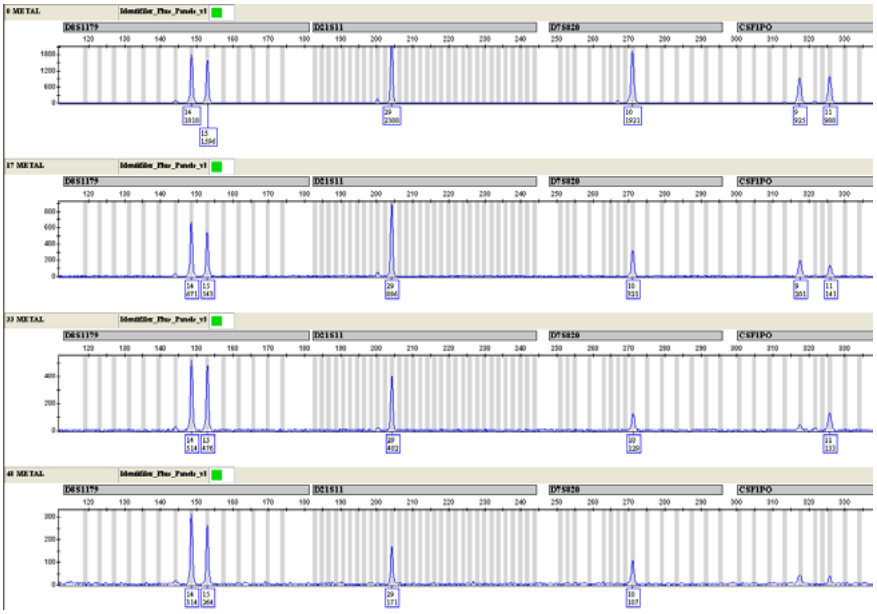


Figure 1. a graph showing a schematic representation of the DNA depletion pattern in relation to ADD

Discussion

Both Hemastix® strips and the Phenolphthalein test consistently provided positive results, with weak positives appearing at 1357 ADD for Hemastix® and 1215.5 ADD for Phenolphthalein. These presumptive tests rely on the peroxidase-like activity of haemoglobin, which degrades to hemichrome when exposed to environmental factors like temperature and humidity.

In contrast, confirmatory tests—RSID and Hexagon® OBTI—use immunological methods to identify specific blood components. The RSID kit targets Glycophorin A, while OBTI detects human haemoglobin. RSID was less effective with degraded samples, showing negative results first for metal (809 ADD on Day 29), then for glass (943 ADD on Day 33) and cloth (1215.5 ADD on Day 41). The OBTI test failed to yield positive results only for metal and cloth before the final sampling day (1357 ADD on Day 45).

Two key factors likely enhance the OBTI kit's performance. First, Glycophorin A, the target for RSID, undergoes significant biophysical changes; red blood cell membranes can deform within five days, while the oxidation of hemoglobin to

hemichrome occurs more slowly, allowing OBTI to better withstand environmental exposure. Second, haemoglobin constitutes 97% of the blood's dry content, making it more abundant than Glycophorin A. This abundance, along with hemoglobin's slow denaturation, supports OBTI's effectiveness for aged and degraded samples.

The experiment's findings demonstrated the substantial impact of environmental insults on both the quantity and quality of DNA, contradicting previous studies that indicated stability until exposure to extreme conditions such as 100% humidity [8] or temperatures exceeding 150 °C [9]. STR analysis revealed that, except for cloth samples, DNA degradation led to partial profiles, unbalanced peaks, and allelic dropout as ADD approached 1425.5, indicating that controlled environmental changes do not accurately model the effects of natural insults on DNA integrity [10, 2].

Alaeddini and co-authors categorized factors affecting DNA degradation into enzymatic and non-enzymatic processes influenced by environmental conditions. Among the endonucleases involved in DNA degradation during cell death, two subgroups are cation-dependent (Mg^{2+} and Ca^{2+}) [11]. Early in the experiment, desiccation and high temperatures likely inactivated these enzymes, allowing non-enzymatic degradation processes—such as oxidation and hydrolytic reactions— in addition to UV irradiation to dominate, resulting in slow and continuous DNA degradation over time [12].

Conclusion

Presumptive tests demonstrated greater resilience to environmental insults than confirmatory tests for blood. When combined with DNA analysis, confirmatory tests correlated more closely with the ability to obtain STR profiles. While cloth offers effective protection against environmental factors, excessive temperatures may have caused immediate enzymatic inactivation, leading to a slower rate of DNA degradation.

Acknowledgments

We would like to thank RAK Police Head Quarters for supporting this research.

Conflict of interest statement

The authors have no conflicts of interest.

References

[1] Abdel Hady, R. H., Thabet, H. Z., Ebrahem, N. E., et al. (2021). Thermal effects on DNA degradation in blood and seminal stains: Forensic view. Academic Forensic Pathology, 11(1), 7-23.

[2] Lund, S., & Dissing, J. 2004, April. Surprising stability of DNA in stains at extreme humidity and temperature. In International Congress Series, 1261, 616-618. Elsevier.

[3] Bogas, V., Carvalho, M., Anjos, M. J., et al. (2016). Genetic identification of degraded and/or inhibited DNA samples. Australian Journal of Forensic Sciences, 48(4), 381-406.

[4] Rahi, G. S., Adams, J. L., Yuan, J., et al. (2021). Whole human blood DNA degradation associated with artificial ultraviolet and solar radiations as a function of exposure time. Forensic Science International, 319, 110674.

[5] Hernández, A., Martínez, R., & Morales, M. (2022). Environmental impacts on DNA quality and recovery from forensic samples: A review. *Forensic Sciences Research*, 7(1), 1-14.

[6] www.timeanddate.com/weather/united-arab-emirates/ras-al-khaimah. Accessed daily from 31st March until 19th May 2013.

[7] Walsh, P.S., Metzger, D.A., & Higuchi, R. 1991. Chelex 100 as a Medium for Simple Extraction of DNA for PCR-Based Typing from Forensic Material. BioTechniques, 10 (4) 506-513.

[8] Dissing, J., Søndervang, A., & Lund, S. 2010. Exploring the limits for the survival of DNA in blood stains. Journal of Forensic and Legal Medicine, 17, (7) 392-396.

[9] Barbaro, A., & Cormaci, P. 2008. Study about the effect of high temperatures on STRs typing. Forensic Science International: Genetics Supplement Series, 1, (1) 92-94.

[10] Kobilinsky, L. 1992. Recovery and stability of DNA in samples of forensic science significance. Forensic Science Review, 4, (1) 68-87.

[11] Alaeddini, R., Walsh, S. J., & Abbas, A. 2010. Forensic implications of genetic analyses from degraded DNA—a review. Forensic Science International-Genetics, 4, (3) 148-157.

[12] Hofreiter, M., Serre, D., Poinar, H. N., et al. 2001. Ancient DNA. Nature Reviews Genetics, 2, (5) 353-359.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 705-712
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Disclosing the maternal genetic background of Andean Mestizos from Ecuador

Alejandra Garzón-Salazar[1,2*], Germán Burgos[3,4], Rodrigo Flores-Espinosa[5], Katherin Barrionuevo-Perez[6,7], Ignacio Yépez[6], Alejandro Cabrera-Andrade[8,9], Elius Paz-Cruz[10,11], Andrés Ordóñez-Ugalde[12,13,14], António Amorim[1,2,15], Leonor Gusmão[5], Filipa Simão[16], Verónica Gomes[1,15]

Filipa Simão and Verónica Gomes should be considered joint senior author.

[1] *i3S- Instituto de Investigação e Inovação em Saúde, Porto, Portugal*

[2] *Biology Department, Faculty of Sciences, University of Porto, Porto, Portugal*

[3] *One Health Research Group, Facultad de Medicina, Universidad de las Américas, Quito, Ecuador*

[4] *Grupo de Medicina Xenómica, Universidad de Santiago de Compostela, Santiago de Compostela, España*

[5] *Laboratório de Diagnóstico por DNA, Rio de Janeiro, Brazil*

[6] *Laboratorios de Investigación- Universidad de Las Américas, Quito, Ecuador*

[7] *Carrera de Biotecnología, Universidad de Las Américas, Quito, Ecuador.*

[8] *Grupo de Bio-Quimioinformática, Universidad de Las Américas, Quito, Ecuador*

[9] *Carrera de Enfermería, Facultad de Ciencias de la Salud, Universidad de Las Américas, Quito, Ecuador*

[10] *Laboratorio de ADN, Fiscalía General del Estado, Quito, Ecuador*

[11] *Centro de Investigación Genética y Genómica, Facultad de Ciencias de la Salud Eugenio Espejo, Universidad UTE, Quito, Ecuador*

[12] *Laboratorio Biomolecular, Cuenca, Ecuador*

[13] *Instituto de Investigación Sanitaria y Fundación Pública Galega de Medicina Xenómica*

[14] *Centro de Investigación Biomédica en Red de Enfermedades Raras, Santiago de Compostela, España*

[15] *IPATIMUP – Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal*

[16] *Department of Forensic Science, Virginia Commonwealth University, USA.*

** Corresponding author at: R. Alfredo Allen 208, 4200-135 Porto, ale22gs@outlook.com*

Abstract
Ecuador retains European and African genetic contributions from the colonial period alongside its Native American ancestry. Mestizos, individuals with Native, European, and African inheritance, are currently distributed across various regions of the country. This study assessed the maternal genetic composition of Mestizos from the Ecuadorian Andes to understand the impact of historical migration. Complete mtDNA control region sequences were analyzed from 214 individuals across three subregions: North, Center, and South Andes. High haplotype diversity values were observed in all regions, with Native American haplogroups prevailing across the sample. Differences in the proportions of certain lineages were found among regions. African and Eurasian haplogroups were detected at low frequencies in North and South Andes but were absent in Central Andes, where most Native American communities are established. Despite some variation in haplogroup distribution among Andean sub-regions, there were no significant differences in haplotype composition. In a South American context, our Ecuadorian samples showed genetic similarity to West Andean populations, driven by elevated haplogroup B frequency, suggesting a distinct evolutionary path to the populations from the Eastern South American regions.

Keywords
Ecuador, Andean region, Mestizo, mtDNA, maternal diversity

Introduction

South American human populations experienced a series of complex admixture events that shaped their genetic makeup. The Ecuadorian territory was invaded by the Incas in the XV century and by the Spanish in the XVI century [1]. Currently, the Ecuadorian population is a reflection of admixture processes resulting from the coexistence of several Native ethnic groups, European settlers, and enslaved Africans. According to the last census in 2022, 77.5% of the population self-identified as Mestizos, individuals that retained Native, European, and African genetic inheritances [2]. However, the admixture events occurred differently across the three Ecuadorian regions (Coast, Andes, and Amazon) [3].

Until now, the variability of maternal lineages between and within Ecuadorian regions is unknown, especially in the Andes, which was the home of several Native American groups. In this study, the maternal genetic inheritance of the Andean region of Ecuador was assessed, with the aim of evaluating how historical events have shaped the genetic composition along the Andes and how it may have affected the diversity of the mestizo population in the country.

Material studied, methods, techniques

Population samples and DNA extraction

Blood or mouth swab samples were collected from 214 self-identified Mestizo individuals across three Ecuadorian Andean regions: North (n=76), Center (n=66), and South (n=72). DNA extraction was performed using a salting-out method for blood and phenol-chloroform for swabs. DNA quantification was conducted using a NanoDrop™ 2000/2000c spectrophotometer (Thermo Fisher Scientific).

Mitochondrial DNA analysis

The mtDNA control region (16024-576) was amplified by PCR in a single amplicon. PCR product was confirmed by gel electrophoresis, and products were purified using Alkaline phosphatase (FastAP) and Exonuclease I enzymes (Thermo Fisher Scientific). Cycle sequencing was performed using the Big-Dye Terminator v3.1 kit (Thermo Fisher Scientific). After sequencing, products were purified with Sephadex™ G-50 size exclusion resin at 10% (GE Healthcare). Fragments were separated and detected via capillary electrophoresis on an ABI 3500 Genetic Analyzer (Applied Biosystems®).

Data Analysis

Sequences were aligned with the rCRS [5] using the SeqScape® Software 3 (Applied Biosystems®) to determine haplotypes. Haplogroups were assigned using the EMPOP database, and the Haplosearch tool [6] was employed to convert haplotypes to DNA sequences for further analysis. Arlequin v3.5 [7] was used to calculate genetic diversity parameters and perform analyses of molecular variance (AMOVA). Multi-Dimensional Scaling (MDS) and Principal Component Analysis (PCA) were computed in STATISTICA 10 software (TIBCO Software Inc.).

Results

a. Genetic Diversity and Haplogroup Composition

High haplotype diversity values were found in the three Andean regions studied (>0.9960). The majority of the mtDNA haplogroups found in our sample were of Native American origin (A2, B2, B4, C1, D1, and D4) with differences in the proportion of some lineages among regions (*Figure 1*). African (L1, L2, L3) and Eurasian haplogroups (U4c1, HV, R) were only encountered at low frequencies in North and South Andean regions.

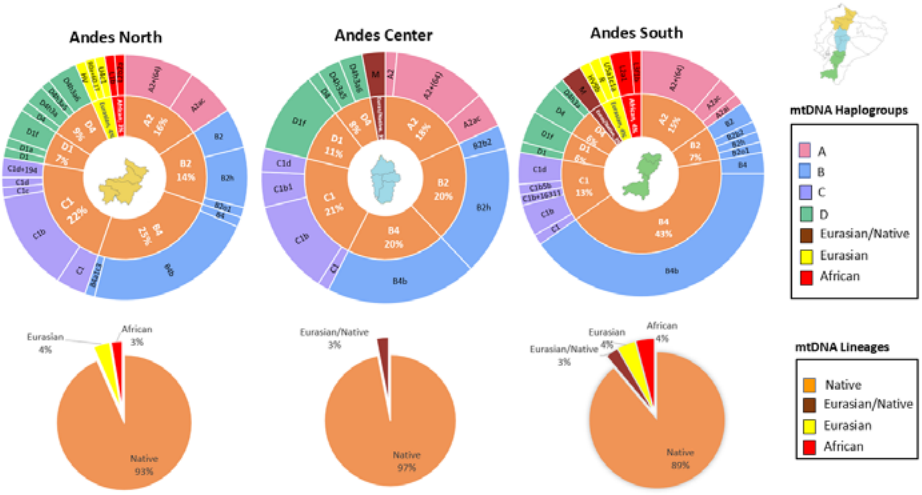


Figure 1. Representation of mtDNA haplogroups and sub-haplogroups found in each Andean population from Ecuador.

b. Population Structure

Although some differences can be detected between Andean populations based on haplogroup frequencies (*Figure1*), pairwise F_{ST} genetic distances based on haplotype frequencies showed no statistically significant differences (data not shown). Analysis of molecular variance showed that 99.73% of the variation is encountered within populations and only 0.27% is attributed to differences between populations (P=0.22535±0.00411).

c. Ecuadorian Andes in a South American Context

An analysis of pairwise genetic distances was carried out to assess the impact of geography on the genetic composition of South American populations. In addition, a map of haplogroup distributions was also created to correlate genetic distances and ancestry proportions.

Although displaying similarities regarding major maternal Native American heritage and minor African and Eurasian contributions, Western South American populations showed some genetic differentiation in the maternal background, which is clear by the absence of a well-defined cluster in the MDS plot (*Figure 2A*). Contrarily, similar haplogroup distributions were observed among Andean Ecuadorean and Peruvian populations (*Figure 2B*).

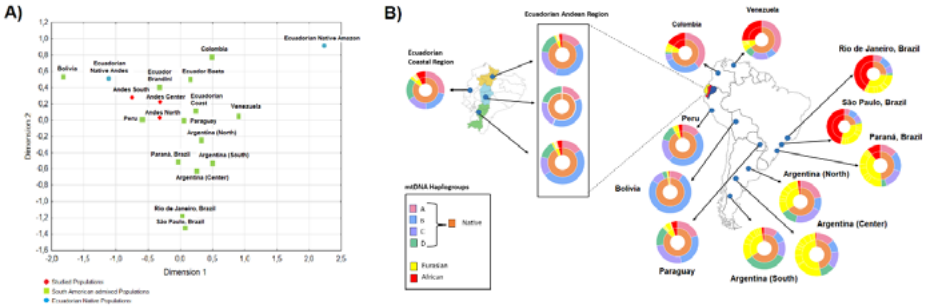


Figure 2. A) MDS plot based on haplotype *FST* genetic distances (Stress=0.08). B) Maternal lineages found in several admixed populations from South America.

To gain a deeper understanding of the genetic structure and relationships among these populations, new analyses were performed by focusing exclusively on the mtDNA lineages of Native American origin (A, B, C, D), excluding all lineages of European or African origin. A PCA was performed using only the Native component to identify the haplogroups that played a role in the differentiation of South American populations.

The PCA (data not shown) revealed clear differences between populations from the Western and Eastern Andes, particularly marked by the higher frequency of haplogroup B in the Western Andes.

Discussion

High haplotype diversity values are expected in admixed populations. However, our studied sample showed a value higher than reported for neighboring countries such as Colombia (0.9932 ± 0.0039) [8]. The prevalence of Native American lineages in the Central Andes was expected, considering that it is in this area that most Native American communities are established. African haplogroups detected in Andes can most likely be traced to the colonial period [9,10], although their presence due to more recent migrations cannot be discarded, especially in the South where evidence of African settlements is lacking. Similarly, deciphering the time of origin of Eurasian haplogroups in Andean populations is a challenging task, as they may be remnants from colonial times, Jewish migrations during World War II, or more recent migrations [11].

The absence of discernible differences among the three Andean populations may be attributed to gene flow that occurred over time. Factors such as intermarriage and domestic migration could contribute to a genetic homogenization of the populations, potentially erasing distinct genetic signatures that might have once characterized specific Native groups. Moreover, despite the occurrence of non-Native haplogroups in different proportions among populations in the Northern and Southern regions, and their absence in the central region (Figure 1), the differences observed are not high enough to detect a statistically significant differentiation between three regions ($P \geq 0.05$).

Considering all South America, Western populations exhibited genetic differentiation primarily due to variations in the proportions of specific Native lineages. For instance, populations in the Northwestern regions of Colombia and Venezuela display a greater genetic distance from Peruvian or Ecuadorian populations (Figure 2), likely due to the high frequency of the Native haplogroup A, in contrast to the prevalence of the Native haplogroup B in Central West Andean populations [12]. The genetic proximity between Ecuadorean Andes and Peru might be attributed to historical population interactions facilitated by their geographical proximity [13]. The predominance of haplogroup B in Andean populations could be traced back to the earliest Native inhabitants. Andean nations were part of the Inca Empire, whose rapid territorial expansion and population growth led to the disappearance of smaller population groups. Rodriguez-Delfin et al. (2001) suggest that the distribution of haplogroup B in these populations could be explained by its predominance in a possible small initial Inca population. According to these authors, genetic drift could have been the cause of the prevalence of this haplogroup, and its current distribution pattern observed across the Andes [12,14,15,16].

Conclusion

Andean populations in Ecuador showed a strong preservation of Native American genetic inheritance, with minimal European and African maternal lineages, in contrast to regions in Eastern South America. The absence of statistically significant differences among the Andean regions indicates limited maternal population substructure. The genetic variations in mtDNA suggest a unique evolutionary history and genetic composition for Ecuadorian Andean populations, highlighting complex ancestral contributions and supporting a separate evolutionary path compared to their Eastern South American counterparts.

Acknowledgments

IPATIMUP integrates the i3S research unit, which is partially supported by the Portuguese Foundation for Science and Technology (FCT). VG is supported by FCT, under the program contract provided in Decree-Law no.57/2016 of August 29. GB and KB-P are supported by MED.GBF.22.03 grant from Dirección General de Investigación, Universidad de Las Américas, Quito, Ecuador.

Conflict of interest statement

None.

References

[1] F. M. Salzano and M. C. Bortolini, "The Evolution and Genetics of Latin American Populations," *Journal of Heredity*, vol. 93, no. 5, pp. 387–387, Sep. 2002, doi.org/10.1093/JHERED/93.5.387.

[2] INEC (Instituto Nacional de Estadisticas y Censos), "Resultados del Censo 2022," 2022, [Online]. Available: http://www.ecuadorencifras.gob.ec/resultados/

[3] E. Ayala, *Resumen de Historia del Ecuador*. 2003.

[4] F. Sanger, S. Nicklen, and A. R. Coulson, "DNA sequencing with chain-terminating inhibitors," *Proc Natl Acad Sci U S A*, vol. 74, no. 12, pp. 5463–5467, 1977, doi.org/10.1073/PNAS.74.12.5463.

[5] R. M. Andrews, I. Kubacka, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull, and N. Howell, "Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA," *Nature Genetics 1999 23:2*, vol. 23, no. 2, pp. 147–147, 1999, doi.org/10.1038/13779.

[6] R. Fregel and S. Delgado, "HaploSearch: A tool for haplotype-sequence two-way transformation," *Mitochondrion*, vol. 11, no. 2, pp. 366–367, Mar. 2011, doi.org/10.1016/J.MITO.2010.11.001.

[7] L. Excoffier and H. E. L. Lischer, "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows," *Mol Ecol Resour*, vol. 10, no. 3, pp. 564–567, May 2010, doi.org/10.1111/J.1755-0998.2010.02847.X.

[8] A. Castillo, L. Deccache, E. F. de Carvalho, F. Simão, and L. Gusmão, "Maternal genetic characterization of a Colombian Andean population," *Forensic Sci Int Genet Suppl Ser*, vol. 7, no. 1, pp. 342–344, Dec. 2019, doi.org/10.1016/J.FSIGSS.2019.10.004.

[9] A. Moya Torres, "Ethnos Atlas etnográfico del Ecuador," 1997, Accessed: Mar. 17, 2023. [Online]. Available: http://repositoriointerculturalidad.ec/jspui/handle/123456789/33472

[10] Lola. Vásquez S., N. Saltos Galarza, and Fundación "José Peralta.," "Ecuador : su realidad," p. 420, 2013.

[11] E. Gil Blanco and A. J. Canela Ruano, "La migración europea al Ecuador (1935-1955). La visión diplomática de la llegada de colectivos problemáticos: los judíos," 2018, Accessed: Jun. 25, 2023. [Online]. Available: http://repositorio.flacsoandes.edu.ec/handle/10469/15780

[12] A. Gómez-Carballa *et al.*, "The peopling of South America and the trans-Andean gene flow of the first settlers," *Genome Res*, vol. 28, no. 6, pp. 767–779, Jun. 2018, doi.org/10.1101/GR.234674.118/-/DC1.

[13] F. Blum, C. Barrientos, A. Ingunza, and Z. Poirier, "A phylolinguistic classification of the Quechua language family," *Indiana*, pp. 1–20, 2023, doi.org/10.31235/OSF.IO/TWU6A.

[14] J. R. Sandoval *et al.*, "The Genetic History of Indigenous Populations of the Peruvian and Bolivian Altiplano: The Legacy of the Uros," *PLoS One*, vol. 8, no. 9, p. 73006, 2013, doi.org/10.1371/JOURNAL.PONE.0073006.

[15] W. S. Watkins *et al.*, "Genetic analysis of ancestry, admixture and selection in Bolivian and Totonac populations of the New World," *BMC Genet*, vol. 13, no. 1, pp. 1–14, May 2012, doi.org/10.1186/1471-2156-13-39/TABLES/2.

[16] L. A. Rodriguez-Delfin, V. E. Rubin-de-Celis, and M. A. Zago, "Genetic diversity in an Andean population from Peru and regional migration patterns of Amerindians in South America: data from Y chromosome and mitochondrial DNA," *Hum Hered*, vol. 51, no. 1–2, pp. 97–106, 2001, doi.org/10.1159/000022964.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 713-722
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Expanding population data for mtDNA haplogroups in the Republic of North Macedonia by using mtDNA set of SNPs in SNP microarray analysis

Renata Jankova[1], Robert Janevski[2], Meri Kokich[2], Sara Kocevska[2], Ognen Boshkovski[2]

[1]*Institute of Forensic Medicine, criminology and medical deontology, Faculty of Medicine, University Ss."Cyril and Methodius", Skopje, R. North Macedonia*

[2]*G-life, Skopje, R. North Macedonia*

Corresponding author at: Institute of Forensic medicine, criminology and medical deontology, Faculty of Medicine, Majka Tereza 19, 1000 Skopje, R. North Macedonia

E-mail address: renjankova@yahoo.com

Abstract

The analysis of mtDNA SNPs has become a widely used method in disease analysis, forensic genetics, molecular anthropology, but also in population genetic studies. While routine procedures for mtDNA analysis is by sequencing the control region, in recent years there have also been used several other techniques to analyse single nucleotide polymorphisms (SNPs) on the mtDNA genome, for example pyrosequencing and the most common used method SNaPshot techniques. Mitochondrial DNA haplogroups are mainly discriminated by coding region SNPs, but covering the whole mtDNA genome provides more comprehensive understanding of the genetic variation including subclades within major haplogroups.

Here we present the results of the analysis of 1765 mtDNA SNPs throughout the whole mitochondrial genome which are part of SNP microarray - Global Screening Array (GSA) v3 microarray, (Illumina, USA), that also covers nuclear DNA SNPs which primarily are used for detection of variants in different health conditions.

The size of the sample group for determination of the mtDNA haplogroup frequencies was 100 individuals from the Macedonian ethnic group from the Republic of North Macedonia. Raw data from the SNParray were analyzed by Plink v1.90b7.2. The mitochondrial sequences were imported in the mthap software Data Version 17.9 – (2016-02-18), and annotated variants were imported in EMPOP database (v4/R13) for haplogroup identification.

The obtained results were in concordance with previously published data for mtDNA haplogroups in the Republic of Macedonia. In our study the most common haplogroup was H with 37%, followed by haplogroup U with 24%, and intermediate frequencies of haplogroups J (9%),

K (7%), N (5%) and T (4%). Within haplogroups, subclades were determined among which 3 subclades from the L macrohaplogroup. Microarray SNPs covering the whole mtDNA genome allow determination of the subclades of haplogroups and also give a possibility to enrich the population data of the country with additional information.

Keywords
mtDNA SNPs, microarray, haplogroups, population study

1. Introduction

Except for the Sanger sequencing of the control region [1], there have recently been used several other techniques for mtDNA sequence analysis such as analyses of single nucleotide polymorphisms (SNPs) on the whole mtDNA genome using SNaPshot techniques [2] or microarray system by pyrosequencing [3]. Mitochondrial DNA haplogroups are mainly determined by coding region SNPs, but covering the whole mtDNA genome provides a more comprehensive understanding of the genetic variation including subclades within major haplogroups. The mtDNA SNPs analysis has become a widely used method in disease analysis [4], forensic genetics [5], molecular anthropology [6], and in population genetic studies [7,8]. Mitochondrial DNA haplogroups are mainly discriminated by coding region SNPs, but covering whole mtDNA genome provides more comprehensive understanding of the genetic variation allowing determination of the subclades within major haplogroups.

2. Materials and Methodes

2.1. Sample collection and preparation

The size of the sample group for determination of the mtDNA haplogroup frequencies was 100 individuals of Macedonian ethnicity from North Macedonia. Consent was obtained from all 100 individuals for genetic testing after which a buccal swab was taken and analysed. DNA was extracted using the QIAamp mini kit (Qiagen, Hilden, Germany) following the manufacturer's recommendations. Samples were prepared for SNP genotyping using Illumina's Infinium HTS assay reference guide and manual protocol on the Infinium Global Screening Array-24 BeadChip, v3.

2.2. Analysis of mtDNA SNPs

We analyzed 1765 mtDNA SNPs from the whole mitochondrial genome, per sample, which are part of SNP microarray - Global Screening Array BeadChips that contains 700 000 SNPs (Illumina, USA).

Analyses were performed on Illumina Infinium using Illumina global screening array BeadChips and analysed with Illumina iScan System using GenomeStudio® 2.0 software [9].

Raw data was analyzed by Plink software v1.90b7.2. [10]. The mitochondrial DNA sequence was imported in the Mthap software Data Version 17.9 – (2016-02-18) for automatic variants detection according to the PhyloTree build 17 [11], and annotated variants were imported in EMPOP database (v4/R13) for haplogroup identification [12,13].

3. Results

The obtained results of the polymorphic sites of mtDNA in 100 unrelated Macedonian individuals and established mtDNA haplogroups showed that haplogroup cluster HV is the most common with 39%, including haplogroups HV0 (2%) and haplogroup H with its subhaplogroups (37%). Haplogroups observed at intermediate frequencies included clusters U (24%), J (9%), K (7%), N (5%) and T (4%). The haplogroups observed less frequently have included haplogroup L (3%) and haplogroups R, V, W and X with 2% prevalence. Observed polymorphic sites in analyzed 100 unrelated Macedonian individuals and established mtDNA haplogroups are presented in table 1. Percentages of these haplogroups are presented in figure 1.

Table 1. Identified polymorphisms and determined mtDNA haplogroups in 100 samples

Sample number	Haplogroup	Genotype
428665MT	D4j8	C5178a T16362C G3010A C8414T C14668T G11696A C16174T
247959MT	H12a	G2706A T7028C C3936T C16287T
328574MT	H	G2706A T7028C
892620-MT	H	G2706A T7028C
904312-MT	H	G2706A T7028C
775276-MT	H	G2706A T7028C
351130MT	H12a	G2706A T7028C C3936T C16287T

389881MT	H	G2706A T7028C
530223MT	H	G2706A T7028C
568889MT	H	G2706A T7028C
597211MT	H15	G2706A T7028C T6253C
466297MT	H1	G2706A T7028C G3010A
511089MT	H1	G2706A T7028C G3010A
963782-MT	H11	G2706A T7028C T8448C G13759A T16311C!
433367MT	H13b1	G2706A T7028C C14872T T11485C C16261T A200G
314642MT	H17	G2706A T7028C G3915A
466479MT	H17	G2706A T7028C G3915A
263673MT	H1b	G2706A T7028C G3010A T16356C
831640-MT	H1b1	G2706A T7028C G3010A T16356C A3796G
520897MT	H1c	G2706A T7028C G3010A T477C
255568MT	H1e	G2706A T7028C G3010A G5460A
353707MT	H1t	G2706A T7028C G3010A G9986A
862509-MT	H1u2	G2706A T7028C G3010A C9923T G3483A C16320T
528844MT	H12	G2706A T7028C C3936T
186354MT	H2a2a1	G2706A T7028C G1438A G4769A G750A G8860A G15326A
201998	H2a2a1	G2706A T7028C G1438A G4769A G750A G8860A G15326A
133673MT	H3	G2706A T7028C T6776C
232974MT	H3	G2706A T7028C T6776C
755384-MT	H4	G2706A T7028C C3992T T5004C G9123A
348237MT	H48	G2706A T7028C G11016A
204114	H5	G2706A T7028C T16304C
201900116-MT	H55	G2706A T7028C G10646A
187896	H55	G2706A T7028C G10646A
162636MT	H55b	G2706A T7028C G10646A T16189C!
238918MT	H55b	G2706A T7028C G10646A T16189C!
261338MT	H55b	G2706A T7028C G10646A T16189C!
241015MT	H5b	G2706A T7028C T16304C G5471A
959248-MT	H5b	G2706A T7028C T16304C G5471A
201900107-MT	HV	T14766C

195329MT	HV0	T14766C T72C T16298C
817489-MT	J1b1a1	A10398G! A12612G G13708A C16069T C462T G3010A G8269A G16145A C16261T G5460A T13879C
140857MT	J1c	T489C A10398G! A12612G G13708A C16069T C462T G3010A T14798C
186603MT	J1c	T489C A10398G! A12612G G13708A C16069T C462T G3010A T14798C
285804	J1c	T489C A10398G! A12612G G13708A C16069T C462T G3010A T14798C
189935MT	J1c2	T489C A10398G! A12612G G13708A C16069T C462T G3010A T14798C A188G
702478-MT	J1c2	T489C A10398G! A12612G G13708A C16069T C462T G3010A T14798C A188G
360325MT	J1c2	T489C A10398G! A12612G G13708A C16069T C462T G3010A T14798C A188G
664239MT	J1c2e	T489C A10398G! A12612G G13708A C16069T C462T G3010A T14798C A188G C16366T
664269	J1c2e	T489C A10398G! A12612G G13708A C16069T C462T G3010A T14798C A188G C16366T
798238-MT	K1	A10550G T11299C T14798C T16224C T16311C! T1189C A10398G!
826456-MT	K1a	A10550G T11299C T14798C T16224C T16311C! T1189C A10398G! C497T
627088MT	K1a15	A10550G T11299C T14798C T16224C T16311C! T1189C A10398G! C497T G8155A
407952MT	K1a4	A10550G T11299C T14798C T16224C T16311C! T1189C A10398G! C497T T11485C
661601MT	K1c1	A10550G T11299C T14798C T16224C T16311C! T1189C A10398G! T146C! T152C! C498d A9093G G11377A
612604MT	K1c2	A10550G T11299C T14798C T16224C T16311C! T1189C A10398G! T146C! T152C! C498d A14002G G14040A C16320T
291050MT	K2b1a1	A10550G T11299C T14798C T16224C T16311C! T146C! T9716C C2217T G5231A C11869a G13135A C16270T
590374MT	L0a1a	G263A C146T G5231A G5460A C16320T T3866C
510472MT	L1b1a16	G3666A A7055G T7389C T13789C T14178C G14560A G185t A357G T710C G1438A T1738C T2352C A2768G T3308C G3693A C6548T T6827C C7867T T12519C A14769G T15115C T16126C A16129G C16270T A5036G G5046A C13880a A14203G T5393C
300656MT	L2a1	T150C! T7175C T182C! A12693G T15784C A16309G
277648MT	N1a	C9540T G10398A A15301G! T10238C G12501A T204C A13780G
559389MT	N1a1	C9540T A15301G! T10238C G12501A T204C A13780G A10398G! G15043A
239800MT	N1b	C9540T G10398A A15301G! G1719A T10238C G12501A G1598A C2639T G5471A G8251A G16390A
931102-MT	N1b	C9540T G10398A A15301G! T10238C G12501A G1598A C2639T G5471A G8251A C16176g G16390A
499034	N1b1	C9540T G10398A A15301G! T10238C G12501A G1598A C2639T G5471A G8251A C16176g G16390A C1703T C3921a C8472T A12822G G16145A
144783MT	R0	T12705C T16223C G73A A11719G
201900113-MT	R0	T12705C T16223C G73A A11719G

666419MT	T2	G709A G1888A A4917G T10463C A15607G G15928A A11812G A14233G
146358MT	T2b	G709A G1888A A4917G T10463C A15607G G15928A C16294T A11812G A14233G C16296T G930A G5147A T16304C
184987MT	T2b	G709A G1888A A4917G T10463C A15607G G15928A C16294T A11812G A14233G C16296T G930A G5147A T16304C
639822MT	T2f	G709A G1888A A4917G T10463C A15607G G15928A C16294T A11812G A14233G C8270T 8281-8289d
269102MT	U1a	A11467G A12308G G12372A C285T T12879C A13104G G15148A A15954c T16249C C2218T G14364A T16189C!
181283-MT	U1a	A11467G A12308G G12372A C285T T12879C A13104G G15148A A15954c T16249C C2218T G14364A T16189C!
121237MT	U2	A11467G A12308G G12372A A16051G
704159-MT	U2e1	A11467G A12308G G12372A A16051G A508G A3720G A5390G T5426C C6045T A10876G T13020C A15907G G16129c T16189C! C340T
267299MT	U2e2	A11467G A12308G G12372A A16051G A508G A3720G A5390G T5426C C6045T T6152C A10876G T13020C T13734C A15907G G16129C T16189C! T16362C T8473C
390678MT	U2e2	A11467G A12308G G12372A A16051G A508G A3720G A5390G T5426C C6045T T6152C A10876G T13020C T13734C A15907G G16129C T16189C! T16362C T8473C
224404MT	U3a1a	A11467G A12308G G12372A C150T A14139G T15454C A16343G C6518T A10506G C13934T G16390A G3010A G7521A!
484818MT	U4	A11467G A12308G G12372A T4646C A6047G T15693C T16356C
882601-MT	U4a	A11467G A12308G G12372A T4646C T15693C T16356C C8818T
331536MT	u4a1	A11467G A12308G G12372A T4646C A6047G T15693C T16356C C8818T T152C! A12937G C16134T
297173MT	U4c1	A11467G A12308G G12372A T4646C A6047G T15693C T16356C A4811G T9070g T11009C C16179T
710045-MT	U4c1	A11467G A12308G G12372A T4646C A6047G T15693C T16356C A4811G T9070g T11009C C16179T
298801MT	U5a1	A11467G A12308G G12372A C16270T A14793G A15218G A16399G
180291MT	U5a1a1	A11467G A12308G G12372A C16270T A14793G A15218G A16399G T1700C T5495C A15924G
995658-MT	U5a1a2a	A11467G A12308G G12372A C16270T A14793G A15218G A16399G C12346T A5319G T6719C
802900-MT	U5a1a2a	A11467G A12308G G12372A C16270T A14793G A15218G A16399G C12346T A5319G T6719C
496901MT	U5a1a2a	A11467G A12308G G12372A C16270T A14793G A15218G A16399G C12346T A5319G T6719C
360272MT	U5a1b	A11467G A12308G G12372A C16192T C16270T A14793G C16256T A15218G A16399G
776630-MT	U5a2a1b	A11467G A12308G G12372A C16270T A14793G G16526A G13928c T13015C
219605MT	U5b	A11467G A12308G G12372A C16270T C150T A7768G T14182C
537478MT	U5b	A11467G A12308G G12372A C16270T C150T A7768G T14182C
572530MT	U5b	A11467G A12308G G12372A C16270T C150T A7768G T14182C

733821-MT	U5b2a	A11467G A12308G G12372A C16270T C150T A7768G T14182C C1721T A13637G
202770MT	U8b1	A11467G A12308G G12372A A1811G A3480G C14167T T16189C! C16234T
989475-MT	V1	G4580A, A8869G
694797-MT	V1	G4580A, A8869G
673218MT	W1e1a	T195C! T204C T1243C A3505G G5460A A11947G C16292T C7864T A8659G C16295T
791860-MT	W3	T195C! T204C T1243C A3505G G5460A A11947G C16292T T1406C
788524-MT	X2b	T6221C C6371T A13966G T14470C T16189C! C16278T! G1719A C8393T G15927A
414048MT	X2m'n	T6221C C6371T A13966G T14470C T16189C! C16278T! G1719A T226C

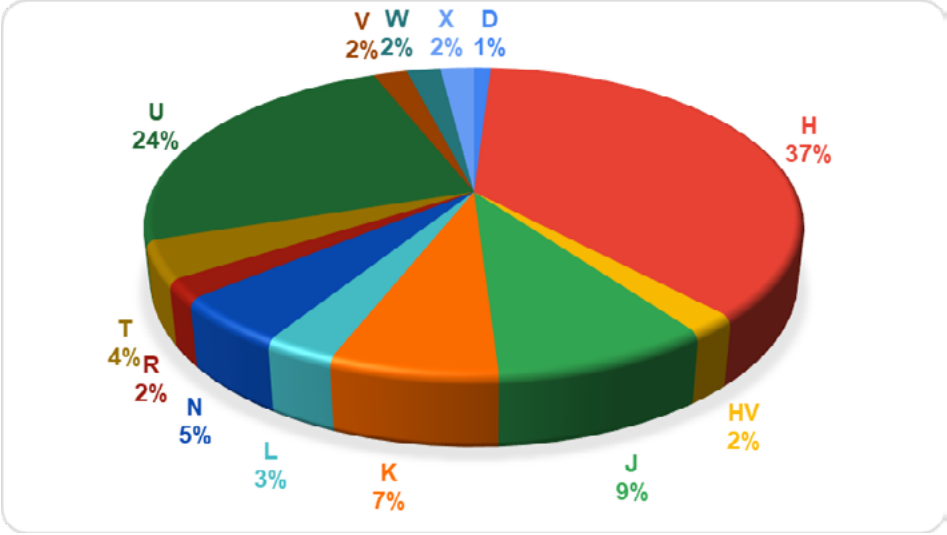


Figure 1. Pie chart showing the percentage of major haplogroups identified within the analyzed individuals.

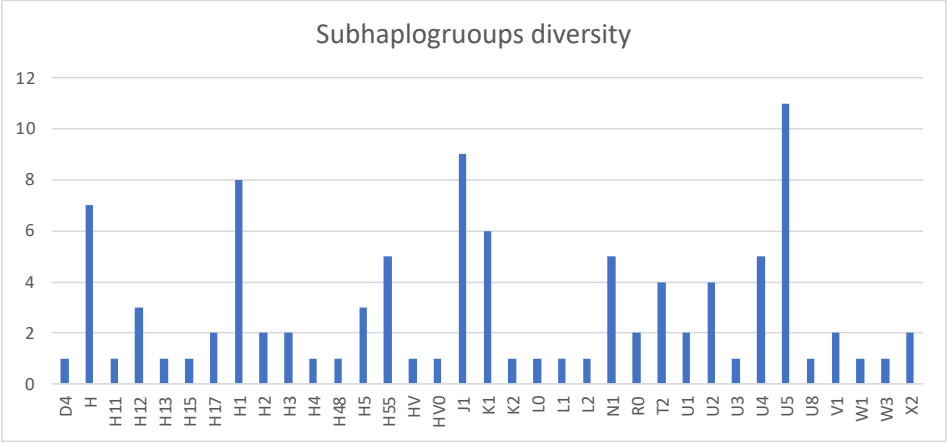


Figure 2. shows the distribution of all established mtDNA subhaplogroups.

Figure 2 Bar chart showing subhaplogroups diversity. Bars below are the established subhaplogroups, whereas numbers on the left are percentages of their presence

As shown in figure 2, except for diversity in H subhaplogroups, high diversity of U haplogroup is observed in Macedonian samples. There are 6 determined U subhaplogroups in ethnic Macedonians out of 9 known U haplogroups. Except for haplogroup I, eight European haplogroups are present in Macedonian population among nine major European haplogroups.

4. Discussion

As expected, the obtained results are in concordance with previously published data for mtDNA haplogroup frequencies in Macedonian population [14]. Also, the basic phylogenetic structure of presented Macedonian sample is in accordance with other West-Eurasian populations.

The presence of three African subhaplogroups L0, L1 and L2 was less expected, even though there is previously published data about presence of L0a haplogroup in one orthodox Macedonian individual [15]. Haplogroup L is generally found with low frequency in Europe, but there are some varieties in different countries. Presence of African L subhaplogroup in the territory of south-eastern European country such as Macedonia could be connected to recent historical events as much as ancient events. Various historical interactions can be explained with ancient human migrations from Africa to Europe including regions that are now

part of North Macedonia, but also modern migrations in more recent times, including movements during and after the World Wars, which could have contributed to the introduction of African mtDNA haplogroups into North Macedonia. Also, the period of Ottoman Empire which controlled the Balkans from 14th century until the early 20th century, included territories in North Africa that might led to inflow of African genetic markers into southeastern Europe.

High diversity of mtDNA U haplogroup in the territory of North Macedonia can be explained by combination of geographical, historical and demographic factors that have shaped the genetic characteristics of the region. The area of North Macedonia as part of the Balkan Peninsula has been a crossroads for various human populations and has been settled since prehistoric times, allowing for the establishment of diverse mitochondrial lineages including those from the U haplogroup. Later on, the incoming agriculturalists likely contributed to this diversity. Also, historical events such as war, plagues and migrations could have led to population bottlenecks, influencing genetic diversity.

5. Conclusion

Microarray SNPs analyses covering the whole mtDNA genome allow the determination of the subclades of haplogroups increasing the phylogenetic resolution. Obtained results enrich our country with additional information of the mtDNA population data which is useful in forensic as well as in population studies.

6. Acknowledgements

We are thankful to all participants and colleagues included in the preparation of this study.

7. References

[1] D. Ballard. Analysis of Mitochondrial control region using Sanger sequencing. Methods Mol.Biol. 1420 (2016)143-55

[2] B. Quintáns, V. Alvares-Iglesias, A. Salas, C. Phillips, MV.Lareu, A.Carracedo. Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing Forensic Sci. Int. 140 (2-3) (2004) 251-7

[3] M. Mikkelsen, FH. Rune, AJ Hansen, N.Morling. Massively parallel pyrosequencing of the mitochondrial genome with the 454 methodology in forensic genetics. Forensic Sci Int Genet 12 (2014) 30-37

[4] C. D. Samuels, D.A. Carothers, R.Horton, P.F. Chinnery. The Power to Detect Disease Associations with Mitochondrial DNA Haplogroups Am J Hum Gen 78 (4) (2006) 713-720.

[5] A. Kloss-Brandstätter, T.J. Parsons, W. Parson. MtDNA coding region SNPs for rapid screening and haplogroup identification of forensic samples. Int Congress Series 1261 (2004) 422-424.

[6] S. Köhnemann, H. Pfeiffer. Application of mtDNA SNP analysis in forensic casework. Forensic Sci Int Genet 5(3) (2011) 216-21.

[7] W. Parson, HJ Bandelt. Extended guidelines for mtDNA typing of population data in forensic science. Forensic Sci Int Genet. 1 (1) (2007) 13-19.

[8] S. Sigurðsson, M. Hedman, P.Sistonen, A. Sajantila, AC. Syvanen. A microarray system for genotyping 150 single nucleotide polymorphisms in the coding region of human mitochondrial DNA. Genomics 87 (2006) 534-542

[9] Illumina Global Screening Array (2019). Available at: https://genebygene.com/products/illumina-global-screening-array/.

[10] S. Purcel, B. Neale, K. Todd-Brown et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet 81 (3) (2007) 559-575

[11] M. van Oven, M. Kayser. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30 (2) (2009) 386-94

[12] W. Parson, A. Dur EMPOP-a forensic mtDNA database. Forensic Sci. Int. Genetics 1(2) (2007) 88-92

[13] W. Parson, A. Kloss-Brandstätter, M. Pircher, R. Scheithauer. EMPOP-the EDNAP mtDNA population database concept for a new generation, high-quality mtDNA database. Forensic Sci. Int. Genet 1261(3) (2004) 106-108

[14] B. Zimmermann, A. Kloss-Brandstätter, N. Duftner. et al. Mitochondrial DNA control region population data from Macedonia 1(3-4) (2007) e4-9

[15] R. Jankova-Ajanovska, Z. Jakovski, S. Risteski, H. Jovanovic et al. African L0a mtDNA haplogroup in Republic of Macedonia. Forensic Sci Int Genet Supplement Series 6 (2017) e308-e309

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 723-730
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Comparison of Fetal DNA Extraction Techniques for Paternity and Fetal Sex Determination: Implications for Forensic Genetics

Patricia Guevara-Ramírez[a+], Santiago Cadena-Ullauri[a+], Viviana A. Ruiz-Pozo[a], Rafael Tamayo-Trujillo[a], Elius Paz-Cruz[a], Marcelo López-Carrera[b], Juan José Builes[c,d], Aníbal Gaviria[e], Ana Karina Zambrano[a*]

[a]*Centro de Investigación Genética y Genómica, Facultad de Ciencias de la Salud Eugenio Espejo, Universidad UTE, Quito, Ecuador.*

[b]*Laboratorio de ADN, Fiscalía General del Estado, Quito, Ecuador*

[c]*Laboratorio Genes SAS, Medellín, Colombia.*

[d]*Instituto de Biología, Universidad de Antioquia, Medellín, Colombia.*

[e]*Hemocentro Nacional, Cruz Roja Ecuatoriana, Quito, Ecuador.*

+The authors contributed equally

**Corresponding author at: Mariana de Jesús S/N and Mariscal Sucre. Centro de Investigación Genética y Genómica, Facultad de Ciencias de la Salud Eugenio Espejo, Universidad UTE, Quito, Ecuador,* anazambrano17@hotmail.com*, +593 2 299 0800 ext. 2192.*

Abstract

Introduction: Forensic genetics significantly depends on precise and reliable techniques to determine paternity, which is often crucial in legal proceedings. The prediction of fetal sex also plays a vital role in various forensic investigations, especially in cases involving crimes against women or infants. Cell-free fetal DNA (cffDNA), circulating in maternal blood, provides a non-invasive method for paternity analysis and fetal sex determination. **Objective:** the objective of this study was to evaluate various techniques for fetal DNA extraction, assessing the advantages and limitations of each method and focusing on DNA concentration and quality to ensure the certainty and precision of forensic genetic analyses, particularly in determining paternity and predicting fetal sex. **Material and methods:** the methods evaluated for fetal DNA extraction included the use of Phenol-Chloroform techniques, a Quick-cfDNA™ Serum & Plasma Kit (Zymo Research, USA), and a QIAamp MinElute ccfDNA Mini Kit (Qiagen, Hilden, Germany). **Results and Conclusions:** The study revealed that the QIAamp MinElute ccfDNA Mini Kit exhibited the highest DNA concentration; however, by performing capillary electrophoresis, only the cffDNA extracted using the Quick-cfDNA™ Serum & Plasma Kit was detected. This indicates that the

Quick-cfDNA™ Kit may provide a higher DNA quality or integrity that could be better suited for detection. Therefore, the choice of cffDNA extraction kit should consider not only yield but also the suitability for downstream applications. The concentration of cffDNA in maternal plasma samples was a key consideration, with optimal results obtained using 4 ml of plasma. Plasma processing before DNA extraction was also addressed, emphasizing the importance of immediate processing or appropriate storage conditions, such as -80°C. The concentration of cffDNA in maternal plasma samples, plasma volume, and extraction method emerged as critical factors influencing the quality and viability of extracted DNA. In conclusion, careful attention to these factors is essential for obtaining fetal DNA of sufficient quality and concentration, thereby ensuring the reliability of forensic genetic analyses, particularly in cases involving paternity determination and fetal sex prediction.

Keywords
Forensic genetics, cell-free fetal DNA, Paternity testing, Fetal sex determination.

Introduction

Nowadays, the field of forensic genetics has been increasingly key in legal proceedings, DNA evidence is considered the gold standard for forensic techniques, leading to a greater effect on verdicts in comparison with other non-DNA forensic evidence [1, 2]. For instance, the analysis of short tandem repeats (STRs) is one of the most used techniques for forensic purposes [3]. Furthermore, forensic genetics relies on the accuracy and precision of the techniques employed and its continuous scientific strengthening [4].

DNA paternity testing is decisive in forensic genetics, aiding in the establishment of kinship in legal disputes and by allowing the identification of family members, suspects, victims, or missing persons [5, 6]. Similarly, fetal sex determination could also play a vital role in forensic investigations, given that it can provide key evidence in cases involving crimes against women, and aid in legal disputes over prenatal decisions [7, 8]. Overall, it offers valuable insights into pregnancy-related crimes.

Generally, paternity and fetal sex determination require invasive procedures [7]. Nevertheless, the discovery of cell-free fetal DNA (cffDNA) has allowed the direct analyses of fetal genetic material extracted from a maternal blood sample, this technique holds the potential to refine forensic methodologies [9, 10].

The objective of this study was to evaluate various techniques for fetal DNA extraction, assessing the advantages and disadvantages of each method and focusing on DNA concentration and quality to ensure the certainty and precision of forensic genetic analyses, particularly in determining paternity and predicting fetal sex.

Material studied, methods, techniques

Collection of Maternal Blood and Plasma Preparation

Peripheral maternal blood samples were collected in EDTA tubes from pregnant women within the first 32 weeks of gestation. To maintain sample integrity, processing was completed within 2 hours of collection. Plasma separation began with an initial centrifugation step of 1,900g for 10 minutes at 4°C, which was modified based on the extraction kit methodology. A second centrifugation at 16,000g for 10 minutes was conducted to eliminate any residual cellular debris, ensuring plasma was adequately prepared for subsequent DNA extraction. In cases where immediate processing was not feasible, samples were stored at -80°C to preserve plasma quality.

cffDNA extraction

For the cffDNA extraction, three methods were used. The first method was a modified Phenol-chloroform protocol [11]. The second method was with the Quick-cfDNA™ Serum & Plasma Kit (Zymo Research, USA), and the third method was with the QIAamp MinElute ccfDNA Mini Kit (Qiagen, Hilden, Germany). For the second and third protocols, the manufacturer's recommendations were followed.

cffDNA quantification

For cffDNA concentration quantification, a fluorometry method was used on a Qubit instrument following the manufacturer's protocol.

Capillary electrophoresis

To determine the suitability of the cffDNA for allele detection, capillary electrophoresis was carried out using a 3500 Genetic analyzer with the Y-filer and GlobalFiler PCR amplification kits according to the manufacturer's protocol.

Results

cffDNA concentration

Each cffDNA extraction method was evaluated using various plasma volumes to determine its effectiveness. The cffDNA extraction method using Phenol-chloroform failed to yield results across all tested plasma volumes. In contrast, both the

Quick-cfDNA™ Serum & Plasma Kit and the QIAamp MinElute ccfDNA Mini Kit showed increasing cffDNA concentrations proportional to plasma volumes. Notably, the QIAamp MinElute ccfDNA Mini Kit outperformed the other methods by achieving a higher cffDNA concentration, making it the most efficient cffDNA extraction method in this study (Figure 1).

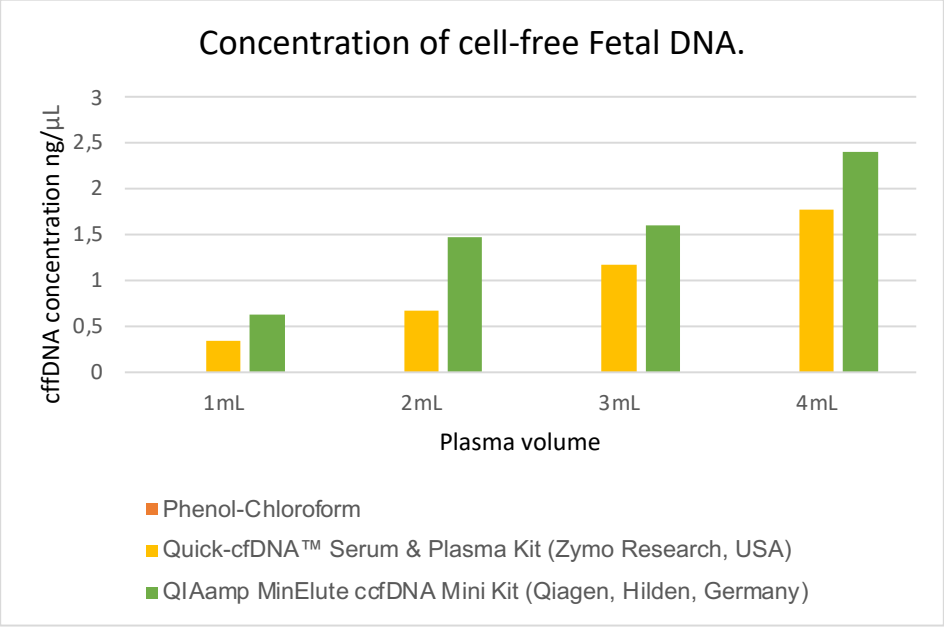


Figure 1. Comparison of cffDNA concentrations using three extraction protocols.

Capillary electrophoresis

During capillary electrophoresis, cffDNA was detected only from the extraction using the Quick-cfDNA™ Serum & Plasma Kit. As shown in Figure 2A, the electropherogram reveals the detection of 23 markers out of 25 (Y-filer). Figure 2B, indicates a mixture of maternal and fetal alleles (GlobalFiler), indicating the presence of both DNA sources.

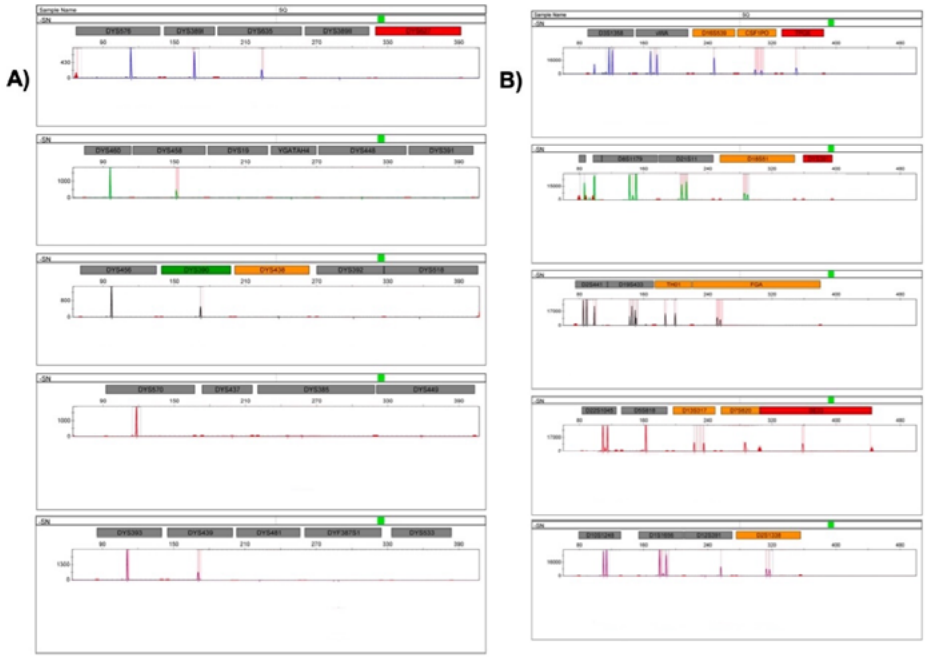


Figure 2. Electropherograms product of the analyses of the extracted cffDNA. A) cffDNA analyzed using the Y-filer PCR amplification kit. B) cffDNA analyzed using the GlobalFiler PCR amplification kit.

Discussion

DNA evidence has been vital in forensic investigations for over three decades. During this period, major technological advancements have been made, enhancing the accuracy, validity of the results, and their applications [12]. Therefore, as this field progresses, developing new technologies remains crucial to enhance forensic analysis.

In this context, the use of cffDNA emerges as a potential new methodology in forensic applications as it can be used as a non-invasive prenatal test for paternity and fetal sex determination, which previously required invasive procedures such as amniocentesis [9].

The phenol-chloroform method is commonly used in DNA extraction [13, 14]. However, this technique has various limitations, including a lower recovery

rate than other cffDNA extraction methods, which is critical when working with cffDNA in maternal blood [15, 16]. In this study, no cffDNA was recovered using the phenol-chloroform extraction method. This result could be associated with the method's low recovery rate, making it unable to isolate cffDNA from the maternal blood samples.

The Quick-cfDNA™ Serum & Plasma Kit and the QIAamp MinElute cffDNA Mini Kit utilize specialized designed columns and reagents for cffDNA extraction, with the latter incorporating magnetic beads for enhanced isolation [17–20]. In the present study, both methods effectively extracted cffDNA but the QIAamp MinElute cffDNA Mini Kit yielded a higher cffDNA concentration, achieving up to 2.5 ng/μL of cffDNA from a starting plasma volume of 4 mL, outperforming the other methods in comparison.

The QIAamp MinElute ccfDNA Mini Kit yielded a higher DNA concentration. However, only the cffDNA extracted using the Quick-cfDNA™ Serum & Plasma Kit was detected via capillary electrophoresis. The authors suggest that the higher cffDNA yield from the QIAamp MinElute Kit may primarily consist of maternal DNA, which could explain the elevated concentration. Since the Y-filer PCR amplification kit exclusively targets Y-STRs, the lower proportion of fetal DNA in the QIAamp MinElute ccfDNA Mini Kit extracts may lead to inadequate Y-chromosome-specific signals, preventing detection. In contrast, the Quick-cfDNA™ Serum & Plasma Kit samples likely contain a higher proportion of cffDNA, facilitating successful Y-STR detection by the Y-filer PCR amplification kit.

These results highlight the importance of selecting an appropriate cffDNA extraction kit, given that the chosen method may not be suitable for specific downstream application.

Conclusion

The Quick-cfDNA™ Serum & Plasma Kit and the QIAamp MinElute cffDNA Mini Kit can effectively extract cell-free fetal DNA. However, despite the QIAamp MinElute ccfDNA Mini Kit yielding higher cffDNA concentrations, only the cffDNA extracted using the Quick-cfDNA™ Serum & Plasma Kit could be detected during capillary electrophoresis. This indicates that the Quick-cfDNA™ Kit may provide a higher DNA quality or integrity that could be better suited for detection. Therefore, the choice of cffDNA extraction kit should consider not only yield but also the suitability for downstream applications.

Acknowledgments

The authors are grateful to Universidad UTE for all their support, and to the Laboratorio de ADN, Fiscalía General del Estado for their collaboration.

Conflict of interest statement

The authors declare no conflicts of interest.

References

[1] Ling S, Kaplan J, Berryessa CM. The importance of forensic evidence for decisions on criminal guilt. *Science and Justice* 2021; 61: 142–149.

[2] Clancy D, Bull R. The Effect on Mock-Juror Decision-Making of Power-of-Speech Within Eyewitness Testimony and Types of Scientific Evidence. *Psychiatry, Psychology and Law* 2015; 22: 425–435.

[3] Zambrano AK, Cadena-Ullauri S, Guevara-Ramírez P, et al. The Autosomal Short Tandem Repeat Polymorphisms Are Potentially Associated with Cardiovascular Disease Predisposition in the Latin American Population: A Mini Review. *Biomed Res Int* 2023; 2023: 1–9.

[4] Arenas M, Pereira F, Oliveira M, et al. Forensic genetics and genomics: Much more than just a human affair. Epub ahead of print 2017. doi.org/10.1371/journal.pgen.1006960.

[5] Molina Y, Raudales I, Moya Z, et al. Characterization of paternity exclusions in the Forensic Medicine service (Honduras) between 2018 and 2022. Epub ahead of print 2022. doi.org/10.2931/RB/2023.08.02.35.

[6] Yudianto A, Setiawan F, Sumino R. Paternity Test Through Kinship Analysis as Forensic Identification Technique. *Majalah Kedokteran Bandung*; 53. Epub ahead of print March 2021. doi.org/10.15395/mkb.v53n1.2154.

[7] Taylor-Sands M, Warton C, Bowman-Smart H. Regulating non-invasive prenatal testing (NIPT) for fetal sex determination. *Med Law Rev* 2023; 31: 521–537.

[8] MEHDER Ö. Fetal sex determination in light of interdisciplinary current studies: a forensic anthropological approach. *Journal of Contemporary Medicine* 2020; 10: 131–137.

[9] Ordoñez E, Rueda L, Cañadas MP, et al. Evaluation of sample stability and automated DNA extraction for fetal sex determination using cell-free fetal DNA in maternal plasma. *Biomed Res Int*; 2013. Epub ahead of print 2013. doi.org/10.1155/2013/195363.

[10] Devaney SA, Palomaki GE, Scott JA, et al. Noninvasive fetal sex determination using cell-free fetal DNA: A systematic review and meta-analysis. *JAMA* 2011; 306: 627–636.

[11] Yuan H, Zhu ZZ, Lu Y, et al. A modified extraction method of circulating free DNA for epidermal growth factor receptor mutation analysis. *Yonsei Med J* 2012; 53: 132–137.

[12] Haddrill PR. Developments in forensic dna analysis. *Emerging Topics in Life Sciences* 2021; 5: 381–393.

[13] Liu AW, Villar-Briones A, Luscombe NM, et al. Automated phenol-chloroform extraction of high molecular weight genomic DNA for use in long-read single-molecule sequencing. *F1000Res*; 11. Epub ahead of print 2022. doi.org/10.12688/f1000research.109251.1.

[14] Corthell JT. DNA and RNA Extraction Protocols. In: *Basic Molecular Protocols in Neuroscience: Tips, Tricks, and Pitfalls*. Elsevier, 2014, pp. 11–19.

[15] Barbier FF, Chabikwa TG, Ahsan MU, et al. A phenol/chloroform-free method to extract nucleic acids from recalcitrant, woody tropical species for gene expression and sequencing. *Plant Methods*; 15. Epub ahead of print 5 June 2019. doi.org/ 10.1186/s13007-019-0447-3.

[16] Madala D, Maktabi MA, Sabbagh R, et al. Lower fetal fraction in clinical cell-free DNA screening results is associated with increased risk of hypertensive disorders of pregnancy. *Prenat Diagn* 2022; 42: 1253–1261.

[17] Diefenbach RJ, Lee JH, Kefford RF, et al. Evaluation of commercial kits for purification of circulating free DNA. *Cancer Genet* 2018; 228–229: 21–27.

[18] Terp SK, Pedersen IS, Stoico MP. Extraction of Cell-Free DNA: Evaluation of Efficiency, Quantity, and Quality. *Journal of Molecular Diagnostics* 2024; 26: 310–319.

[19] Barták BK, Kalmár A, Galamb O, et al. Blood Collection and Cell-Free DNA Isolation Methods Influence the Sensitivity of Liquid Biopsy Analysis for Colorectal Cancer Detection. *Pathology and Oncology Research* 2019; 25: 915–923.

[20] Nesic M, Bødker JS, Terp SK, et al. Optimization of Preanalytical Variables for cfDNA Processing and Detection of ctDNA in Archival Plasma Samples. *Biomed Res Int*; 2021. Epub ahead of print 2021. doi.org/10.1155/2021/5585148.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 731-737
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Genetic ancestry of a Montubios rural population from Ecuador unveils East Asian contribution

Germán Burgos[1,2*], Katherin Barrionuevo-Pérez[3,4], Domenyka Hidalgo-López[5], Andrés Ordóñez-Ugalde [6], Ana Angulo-Pozo[5], Alejandra Garzón-Salazar[3,7,8], Rui Pereira [8,9]

[1] One Health Research Group, Facultad de Medicina, Universidad de Las Américas (UDLA), Quito, Ecuador. [2] Grupo de Medicina Xenómica, Universidad de Santiago de Compostela, Santiago de Compostela, España. [3] Laboratorios de Investigación, Universidad de Las Américas (UDLA), Quito, Ecuador. [3] Laboratorios de Investigación, Universidad de Las Américas (UDLA), Quito, Ecuador. [4] Biotechnology Department, Universitat Politècnica de València, Valencia, Spain. [5] Carrera de Biotecnología, Universidad de Las Américas (UDLA), Quito, Ecuador. [6] Laboratorio Biomolecular Cuenca, Ecuador. [7] Biology Department, Faculty of Sciences, University of Porto, Porto, Portugal. [8] Instituto de Investigação e Inovação em Saúde (i3S), Porto, Portugal. [9] Institute of Molecular Pathology and Immunology, University of Porto (IPATIMUP), Porto, Portugal.

*hgermanburgosf@yahoo.com, german.burgos@udla.edu.ec, hugogerman.burgos@rai.usc.es; *phone +593 995256888; ' current address*

Abstract

The Montubios constitute about 7.4% of the Ecuadorian population. They were officially recognized based on their ethnography in the 2010 census but are between the least studied ethnic groups. Montubios are genetically like the Mestizos group as they also have admixed origin from Native American, European, and African contributors. However, they are settled around the central area of the Pacific coast. In this work, we examine the autosomal genetic ancestry of Montubios rural communities in the Santa Elena province of Ecuador.

53 individuals were analyzed using 46 autosomal ancestry-informative markers (AIM-indels) through multiplex PCR followed by capillary electrophoresis. We then compared Montubio's genetic profiles with four reference populations representing possible main sources of ancestry: African, European, East Asian, and Native American. We performed Bayesian clustering analyses using the STRUCTURE software with the admixture model, to estimate the individual and global ancestry proportions for each population group.

Our results show that the Montubios have a predominantly Native American ancestry (87.1%), followed by European (5.6%), East Asian (5.4%), and African (1.9%) contributions. These findings are consistent with the essentially tri-hybrid admixture model known for most Latin-American

populations, while also showing a noticeable East Asian input in these rural communities from Santa Elena province. Our study is the first to describe the genetic makeup of the Montubios population and contributes to the understanding of the complex genetic diversity of Ecuador and Latin America.

Key Words
Autosomal ancestry, AIMs, rural Montubios, East Asian contribution.

1. Introduction

In Ecuador, as in other regions of Latin America, different historical events occurred that influenced the genetic structure of human populations [1]. The conquest and colonization of the 16th century facilitated the encounter between Native Americans, European settlers, and African slaves, giving rise to an intense process of miscegenation that generated diverse ethnic groups [2], [3]. These groups were officially defined in the 2010 Ecuadorian census as Afro-Ecuadorian, Indigenous, White, Mestizo, and Montubio [4].

The Montubio group is among the least studied and constitutes about 7.4% of the population since their self-recognition based on their ethnography [5]. Although genetically considered like mestizos, as they share African, European, and Native American ancestry, the Montubios settled in southern areas of the Pacific coast [6]. Despite their historical and cultural contribution to the country, information on their ancestry is limited. Therefore, this study proposes an initial approximation to the autosomal ancestry of the rural communities of this ethnic group, which contrasts with urban admixture trends that generally dilute original ancestries.

2. Materials and Methods

2.1. DNA sampling and extraction

53 unrelated self-identified Montubio of rural areas samples were collected in different villages from Santa Elena province, Ecuador. All individuals provided signed informed consent. Saliva samples were extracted using a phenol-chloroform procedure. DNA was then quantified by spectrophotometry (Nanodrop 2000®).

2.2. Amplification and genotyping

PCR amplification of the 46 AIM-InDels was performed through a multiplex reaction [7]. Fragment separation and detection were executed on the ABI 3500

Genetic Analyzer (Applied Biosystems). Genotyping was then assessed with GeneMapper v6 (Applied Biosystems).

2.3. Statistical analysis

Ancestry inferences were performed using STRUCTURE v2.3.4 software [8], from K = 1 up to K = 6, with the Admixture model (while updating allele frequencies using only individuals of putative contributor reference populations with POPFLAG=1 data in an advanced parameter set, namely Native American (n=161), European (n=152), African (n=101) and East Asian (n=229)) to estimate individual and global ancestry proportions for each population group. Each K was tested in six replicate runs consisting of a burn-in length of 100 000 followed by 100000 Markov Chain Monte Carlo (MCMC) interactions. Results were then considered following the best Ln probability of data [7].

4. Results

In this study, we assessed the substructure of the Montubian population with autosomal Ancestry-Informative Markers (AIMs). By showing high divergence between ancestral or geographically distant populations, autosomal AIMs were useful for estimating the distribution of ancestry components of the study population at the continental level [7].

Having explored replicate STRUCTURE analyses from K = 1 up to K = 6, K = 4 emerged as the most appropriate number of clusters in the dataset, considering the plateau reached at that stage for the estimated log-likelihood of the data while corresponding to four human major continental groups.

Our results reveal that Montubios have a predominantly Native American (87.1%) ancestry, followed by European (5.6%), East Asian (5.4%), and African (1.9%) contributions. Additionally, the East Asian ancestry in the population varied across individuals, ranging from 0.2% to 38.4%, with a population mean of 5.4%. This represents a unique ancestral component not previously described in rural Montubio communities.

5. Discussion

These findings were consistent with the essentially trihybrid admixture pattern reported previously. But notably, they present the first report showing a East Asian contribution in a Mestizos-like rural communities in the country.

In Ecuador, genetic analyses using bi-allelic markers such as AIM-InDels or SNPs have been used mainly to study the mestizo population, Afro-Ecuadorians, and specific ethnic groups such as the Kichwas, Shuar, and Cayapas, among others [9], [10], [11]. Previous studies on the Montubio group employing SNPs reported a predominantly Native American contribution followed by European, with a minor African influence, and very low East Asian [4]. Historically, this is consistent with the arrival of Spanish settlers during the conquest of the Native Americans on the Pacific coast, especially in the provinces of Guayas, Manabí, Los Ríos, El Oro, and the southern part of Esmeraldas[6]; bringing African slaves, whose ancestral influence on the Montubio population is observed mainly in the center-north of the pacific coast.

The high percentage of Native American ancestry in rural Montubios is like that of the rest of the country. However, the difference in ancestry concerning urban Montubios is notable, especially because of the significant contribution of East Asian ancestry (0.2% to 38.4% in individuals, with a population mean of 5.4%) when considering four contributing populations (Fig. 1). The ancestral differences observed between rural Montubio, and urban communities may be attributed to the lack of exposure to new migratory events in rural areas. A similar genetic landscape has been observed in other Latin American countries, where native communities, such as the Brazilian Terenas, showed lower genetic diversity concerning urban populations. This reduction in diversity was also observed in the rural population of Santa Isabel, which has been less exposed to European or African influxes than urban populations [12]. Also, in Argentina, studies have indicated a higher presence of European ancestry in urban samples compared to a higher percentage of Native American ancestry in rural samples [13].

The 2022 Ecuadorian Census indicated that 63.1% of the general population is located in urban areas, while in the province of Santa Elena, 43.39% reside in rural areas [5]. These data highlight the demographic, ethnic, and genetic diversity of the country, and emphasize the importance of considering different groups and regions when analyzing the genetic composition of the Ecuadorian population. Studies must not focus exclusively on densely populated urban centers but cover as many populations as possible within a region to ensure representativeness. In this context, rural areas where ancestry is more purely preserved compared to urban counterparts, are crucial for understanding the ancestral origins of a population [14], [15].

Moreover, expanding beyond urban areas is important since mestizos with a higher proportion of indigenous ancestry are found in areas with historically large

native populations, such as the Andean and Mesoamerican regions, while mestizos with higher European ancestry come from areas with low native population density.

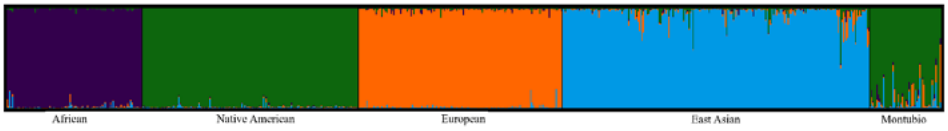


Figure 1. Individual ancestry estimates in ancestral populations (reference) and rural Montubian population (K = 4).

5. Conclusion

The Montubios of rural Ecuador show significant Native American ancestry in their autosomal genetics, reflecting the historical influence of indigenous populations. In addition, they show a notable autosomal contribution of East Asian origin, suggesting historical events of genetic admixture. Accurate data collection in rural areas is essential for obtaining a precise understanding of population demographics and evolutionary history. By intensifying sampling efforts in rural regions of Latin America, we can gain a deeper insight into the genetic diversity and evolutionary dynamics of these under-studied populations.

6. Acknowledgments

To the warm and friendly rural community of the province of Santa Elena. To Carlos Andrés Bastidas-Caldés for his hospitality in San Pedro. To Dr. Manuel Calvopiña for endless hours of information about Ecuador's rural and indigenous populations. This work was funded by DGIV MED.GBF.22.03 grant from Universidad de Las Américas, Quito, Ecuador.

7. Conflict of interest statement

None

8. References

[1] C. Paz-Y-Mino, "La Historia del Ecuador contada por los genes." [Online]. Available: https://www.researchgate.net/publication/287201628

[2] C. M. Bravi, "La evidencia molecular del poblamiento humano de América," *Revista del Museo de La Plata*, vol. 20, no. 177, pp. 40–57, 2013, Accessed: Apr. 15, 2024. [Online]. Available: https://publicaciones.fcnym.unlp.edu.ar/rmlp/article/view/2303

[3] J. C. Long, J. Li, and M. E. Healy, "Human DNA sequences: More variation and less race," *Am J Phys Anthropol*, vol. 139, no. 1, pp. 23–34, May 2009, doi.org/10.1002/AJPA.21011.

[4] S. D. Nagar *et al.*, "Genetic ancestry and ethnic identity in Ecuador," *Human Genetics and Genomics Advances*, vol. 2, no. 4, p. 100050, Oct. 2021, doi.org/10.1016/j.xhgg.2021.100050.

[5] INEC (Instituto Nacional de Estadisticas y Censos), "Resultados del Censo 2022."

[6] Paredes Willington, *Los montubios y nosotros - Willington Paredes - Google Libros*. 2005. Accessed: Apr. 15, 2024. [Online]. Available: https://books.google.com.ec/books/about/Los_montubios_y_nosotros.html?id=vAWzAAAAIAAJ

[7] R. Pereira *et al.*, "Straightforward Inference of Ancestry and Admixture Proportions through Ancestry-Informative Insertion Deletion Multiplexing," *PLoS One*, vol. 7, no. 1, p. e29684, Jan. 2012, doi.org/10.1371/JOURNAL.PONE.0029684.

[8] J. K. Pritchard, M. Stephens, and P. Donnelly, "Inference of Population Structure Using Multilocus Genotype Data," *Genetics*, vol. 155, no. 2, pp. 945–959, Jun. 2000, doi.org/10.1093/GENETICS/155.2.945.

[9] A. K. Zambrano *et al.*, "Ancestry characterization of Ecuador's Highland mestizo population using autosomal AIM-INDELs," 2017, doi.org/10.1016/j.fsigss.2017.09.191.

[10] A. K. Zambrano *et al.*, "The three-hybrid genetic composition of an Ecuadorian population using AIMs-InDels compared with autosomes, mitochondrial DNA and Y chromosome data," *Scientific Reports 2019 9:1*, vol. 9, no. 1, pp. 1–8, Jun. 2019, doi.org/10.1038/s41598-019-45723-w.

[11] R. Santangelo, F. González-Andrade, C. Børsting, A. Torroni, V. Pereira, and N. Morling, "Analysis of ancestry informative markers in three main ethnic groups from Ecuador supports a trihybrid origin of Ecuadorians," *Forensic Sci Int Genet*, vol. 31, pp. 29–33, Nov. 2017, doi.org/10.1016/j.fsigen.2017.08.012.

[12] F. Saloum de Neves Manta *et al.*, "Revisiting the Genetic Ancestry of Brazilians Using Autosomal AIM-Indels," *PLoS One*, vol. 8, no. 9, p. e75145, Sep. 2013, doi.org/10.1371/JOURNAL.PONE.0075145.

[13] M. Caputo, M. A. Amador, A. Sala, A. Riveiro dos Santos, S. Santos, and D. Corach, "Ancestral genetic legacy of the extant population of Argentina as predicted by autosomal and X-chromosomal DIPs," *Molecular Genetics and Genomics*, vol. 296, no. 3, pp. 581–590, May 2021, doi.org/10.1007/S00438-020-01755-W/METRICS.

[14] A. Ruiz-Linares *et al.*, "Admixture in Latin America: Geographic Structure, Phenotypic Diversity and Self-Perception of Ancestry Based on 7,342 Individuals," *PLoS Genet*, vol. 10, no. 9, p. e1004572, Sep. 2014, doi.org/10.1371/JOURNAL.PGEN.1004572.

[15] E. T. Norris *et al.*, "Genetic ancestry, admixture, and health determinants in Latin America," *BMC Genomics*, vol. 19, no. 8, pp. 75–87, Dec. 2018, doi.org/10.1186/S12864-018-5195-7/FIGURES/6.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 739-746
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Analysis of paternity cases with a single exclusion in a genetic marker using Precision ID GlobalFiler™ NGS STR panel v2

Nair Gouveia[1*], Virgínia Lopes[1], Maria João Porto[1], Ana Margarida Bento[1], Lisa Sampaio[1], Armando Serra[1], Filipa Balsa[1], Vanessa Bogas[1], Marta São Bento[1], António Amorim[1,2,3]

[1]*Instituto Nacional de Medicina Legal e Ciências Forenses, Coimbra, Portugal.*
[2]*Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal.*
[3]*REQUIMTE, Analytical Development Group, Laboratório Associado FCT, Lisboa, Portugal.*

** Corresponding author at: nair.b.gouveia@inmlcf.mj.pt*

Abstract

Paternity test results may evidence genetic incompatibilities in the allelic transmission from parents to children, when the expected alleles that the child should have received in a genetic marker do not match to any of the identified alleles in the alleged father and/or biological mother. This exclusion may result from a genetic mutation or other factors as null or silent alleles.

Capillary electrophoresis (CE) is the traditional method used in forensic genetics to analyze STRs (Short Tandem Repeats) for different applications, but Next-Generation Sequencing (NGS) technology may provide additional information, since it allows to detect and sequence simultaneously SNPs (Single Nucleotide Polymorphisms) present in the flanking regions and also distinguish isometric alleles with the same length but different sequences, that were misinterpreted as homozygous.

In this study, a set of reference samples, previously amplified with the GlobalFiler™ PCR Amplification Kit and sequenced by CE on the 3500 Genetic Analyzer, were selected from paternity cases with a single exclusion, reported after GeneMapper ID-X Software analysis. All samples were automatically prepared with the Precision ID GlobalFiler™ NGS STR Panel v2 on the Ion Chef™ System, followed by sequencing on the Ion S5™ System and finally Converge™ Software analysis, according to the manufacturer's instructions.

The aim was to verify if the NGS methodology contributes to identify the parental origin of a mutant allele through sequence data analysis.

The NGS results were in concordance with those obtained by CE. In addition, this methodology demonstrated to be useful to clarify the paternity cases, because it enables a higher power of

discrimination through 9 additional multi-allelic STRs, in a total of 35 markers instead of 24 markers of the GlobalFiler™ PCR Amplification Kit used in the traditional method. Therefore, the Precision ID GlobalFiler™ NGS STR Panel v2 shows to be a powerful method for kinship analyses and typing reference samples.

Keywords
Paternity trios, single exclusion, NGS.

Introduction

In paternity investigations involving the study of the trio composed of the alleged father, mother and child, it is concluded that paternity is excluded only in the presence of 3 or more genetic incompatibilities. However, the analysis of paternity results may reveal the presence of one or even two genetic incompatibilities in the allelic transmission from parents to children, requiring particular attention in the analysis of genetic profiles and also considering the mutation rate of the locus for the statistical calculations [1–3].

Different types of mutations can be observed such as insertion or deletion of a repeat unit and single-base point substitution. Insertion or deletion mutations are usually located in the repeat motif and lead to the detection of the +1 allele or the -1 allele. A point substitution mutation is hardly detected when it occurs in the repeat motif, because it does not lead to a change in the size of the product detected by Capillary electrophoresis (CE), but it is usually detected when it occurs in the primer binding zone. Failure to bind one of the primers means that this allele is not amplified, showing a homozygous genotype of the other allele that was detected. These mutations can result in a discrepancy between the child and one of the parents, due to an alteration in the number of repeats of an allele or they can lead to a lack of sharing of the allele that is not detected in either of the biological parents [1–3].

Null or silent alleles can also be found, which is an allele that is present in the sample, yet is not amplified. A primer-binding site mutation can inhibit amplification for that allele and result in discrepant DNA types at a particular locus [1–3].

The implementation of Next-Generation Sequencing (NGS) technology in the field of forensic genetics may provide additional information when compared to traditional Short Tandem Repeats (STRs) analysis performed by CE. Although the number of genetic markers studied is already high, it is not possible to know the exact allele number variation due to the lack of sequence data and, when necessary,

it is important to consider the study of other genetic markers to investigate the existence of more exclusions. NGS allows to detect and sequence simultaneously SNPs (Single Nucleotide Polymorphisms) present in the flanking regions and also distinguish isometric alleles with the same length but different sequences, that were misinterpreted as homozygous. By identifying repeat motif variation within STRs based on their sequence and not length as in the CE, NGS may be very useful to infer the origin of the genetic mutation and to increase the discrimination power [1–5].

The Precision ID GlobalFiler™ NGS STR Panel v2 is composed by 35 markers, including the same 20 autosomal STRs found in the GlobalFiler™ PCR Amplification Kit used in the CE, along with Y chromosome markers, amelogenin sex marker and additional multiallelic STRs markers [6–7].

The aim of this study was to verify if the NGS methodology using Precision ID GlobalFiler™ NGS STR Panel v2 provides valuable information in paternity cases with a single exclusion in a genetic marker and also to identify the parental origin of a mutant allele.

Material studied, methods, techniques

A total of 45 reference samples (buccal swabs and blood stains), previously amplified with the GlobalFiler™ PCR Amplification Kit and sequenced by CE on the 3500 Genetic Analyzer, were selected from 15 paternity trio cases with a single exclusion, reported after GeneMapper ID-X Software analysis.

As recommended in the user guide, all samples were extracted with the PrepFiler Express™ Kit followed by purification on the AutoMate Express™ System. The extracted samples were quantified with the Quantifiler™ Trio Kit on the 7500 Real-time PCR System and, according to the quantification results, the samples were diluted to 67 pg/μL with nuclease-free water to prepare 15 μL of each diluted sample. Library preparation was automatically prepared with the Precision ID GlobalFiler™ NGS STR Panel v2 using Precision ID DL8 Kit on the Ion Chef™ System. Next, libraries were quantified with the Ion Library TaqMan™ Quantitation Kit on the 7500 Real-time PCR System. All barcoded libraries were diluted to 50 pM and then combined in a super-pool to ensure equal contribution in the sequencing run. Template preparation was performed on the Ion Chef™ System using the Ion S5™ Precision ID Chef & Sequencing Kit and the loaded Ion 530™ chips were sequenced on the Ion S5™ System, as described by manufacturer.

Finally, the results were analyzed with the Converge™ Software and then compared to the results obtained previously with CE.

Results

The NGS results were in concordance with those obtained by CE for all 15 paternity trios (Table 1). For the analysis, it was considered that the allele that the child has in common with the biological parent always comes from maternal origin, as it is the paternity that is being investigated.

Trios 1 to 4 revealed an exclusion at the FGA marker, in which there was a gain or loss of a CTTT repeat in father-child transmission.

Regarding trios 5 and 6, the exclusion was observed in the vWA marker with a mutation resulting from the gain of a TCTA repeat at the final sequence.

In trios 7 and 8, the exclusion occurred at the D12S391 marker. In trio 7, the father's allele 23 mutated to 24 with the gain of an AGAT repeat. For trio 8, the genotype of the child is the same as the mother, which does not allow to distinguish which allele was mutated from the father to the child and there may have been a gain or loss of an AGAT repeat in the initial sequence.

In trio 9, the mutation at the D2S1338 marker was associated with the loss of a TTCC repeat, while for trio 10 with an exclusion at the D7S820 marker, the mutation was the result of the loss of a GATA repeat.

Trio 11 was a particular case, since considering that one of the alleles 13 at the D19S433 marker was necessarily inherited from the mother, there are two possibilities of father-child transmission in which the gain or loss of an AAGG repeat may have occurred.

For trio 12, a mutation was observed in mother-child transmission at the D2S441 marker, in which the child received allele 11 from the father and the allele 12 from the mother mutated to 11 with the loss of a TCTA repeat.

Trios 13 and 14 showed the same flanking SNP rs25768 (A/G) at the D5S818 marker, with the father's allele 12 being mutated to 13 with the gain of an AGAT repeat.

In the case of trio 15 with an exclusion at the CSF1PO marker, a silent or null allele was found, since the allele 8 was not detected in any of the parents in both the CE and NGS methodologies.

Table 1. Summary of NGS results for 15 paternity trios with a single exclusion.

Legend: F – Father; M – Mother; C – Child.

Trio		Allele 1	NGS Sequence Allele 1	Allele 2	NGS Sequence Allele 2
1 FGA	F	23	[TTTC]3 TTTT TTCT **[CTTT]15** CTCC [TTCC]2	25	[TTTC]3 TTTT TTCT [CTTT]17 CTCC [TTCC]2
	M	23	[TTTC]3 TTTT TTCT [CTTT]15 CTCC [TTCC]2	27	[TTTC]3 TTTT TTCT [CTTT]19 CTCC [TTCC]2
	C	22	[TTTC]3 TTTT TTCT **[CTTT]14** CTCC [TTCC]2	23	[TTTC]3 TTTT TTCT [CTTT]15 CTCC [TTCC]2
2 FGA	F	20	[TTTC]3 TTTT TTCT [CTTT]12 CTCC [TTCC]2	26	[TTTC]3 TTTT TTCT **[CTTT]18** CTCC [TTCC]2
	M	22	[TTTC]3 TTTT TTCT [CTTT]14 CTCC [TTCC]2	24	[TTTC]3 TTTT TTCT [CTTT]16 CTCC [TTCC]2
	C	22	[TTTC]3 TTTT TTCT [CTTT]14 CTCC [TTCC]2	27	[TTTC]3 TTTT TTCT **[CTTT]19** CTCC [TTCC]2
3 FGA	F	21	[TTTC]3 TTTT TTCT [CTTT]13 CTCC [TTCC]2	24	[TTTC]3 TTTT TTCT **[CTTT]16** CTCC [TTCC]2
	M	21	[TTTC]3 TTTT TTCT [CTTT]13 CTCC [TTCC]2	23	[TTTC]3 TTTT TTCT [CTTT]15 CTCC [TTCC]2
	C	23	[TTTC]3 TTTT TTCT [CTTT]15 CTCC [TTCC]2	23	[TTTC]3 TTTT TTCT **[CTTT]15** CTCC [TTCC]2
4 FGA	F	19	[TTTC]3 TTTT TTCT **[CTTT]11** CTCC [TTCC]2	25	[TTTC]3 TTTT TTCT [CTTT]17 CTCC [TTCC]2
	M	22	[TTTC]3 TTTT TTCT [CTTT]14 CTCC [TTCC]2	23.2	[TTTC]3 TTTT TT[CTTT]16 CTCC [TTCC]2
	C	20	[TTTC]3 TTTT TTCT **[CTTT]12** CTCC [TTCC]2	22	[TTTC]3 TTTT TTCT [CTTT]14 CTCC [TTCC]2
5 vWA	F	18	[TCTA]1 [TCTG]4 [TCTA]13	18	[TCTA]1 [TCTG]4 **[TCTA]13**
	M	17	[TCTA]1 [TCTG]4 [TCTA]12	18	[TCTA]1 [TCTG]4 [TCTA]13
	C	17	[TCTA]1 [TCTG]4 [TCTA]12	19	[TCTA]1 [TCTG]4 **[TCTA]14**
6 vWA	F	15	[TCTA]1 [TCTG]4 [TCTA]10	18	[TCTA]1 [TCTG]4 **[TCTA]13**
	M	17	[TCTA]1 [TCTG]4 [TCTA]12	19	[TCTA]1 [TCTG]4 [TCTA]14
	C	19	[TCTA]1 [TCTG]4 [TCTA]14	19	[TCTA]1 [TCTG]4 **[TCTA]14**
7 D12S391	F	18	[AGAT]11 [AGAC]6 [AGAT]1	23	**[AGAT]14** [AGAC]9
	M	18	[AGAT]11 [AGAC]6 [AGAT]1	19	[AGAT]12 [AGAC]6 [AGAT]1
	C	19	[AGAT]12 [AGAC]6 [AGAT]1	24	**[AGAT]15** [AGAC]9
8 D12S391	F	17	**[AGAT]10** [AGAC]6 [AGAT]1	21	[AGAT]14 [AGAC]6 [AGAT]1
	M	18	[AGAT]11 [AGAC]6 [AGAT]1	20	[AGAT]13 [AGAC]6 [AGAT]1
	C	18	**[AGAT]11** [AGAC]6 [AGAT]1	20	[AGAT]13 [AGAC]6 [AGAT]1
9 D2S1338	F	23	[TGCC]7 [TTCC]13 GTCC[TTCC]2	26	[TGCC]8 **[TTCC]15** GTCC[TTCC]2
	M	17	[TGCC]5 [TTCC]12	25	[TGCC]7 [TTCC]15 GTCC[TTCC]2
	C	17	[TGCC]5 [TTCC]12	25	[TGCC]8 **[TTCC]14** GTCC[TTCC]2

10 D7S820	F	10	[GATA]10	13	**[GATA]13**
	M	11	[GATA]11	11	[GATA]11
	C	11	[GATA]11	12	**[GATA]12**
11 D19S433	F	12	AAGGTAGG [AAGG]10	14	AAGGTAGG **[AAGG]12**
	M	13	AAGGTAGG [AAGG]11	15	AAGGTAGG [AAGG]13
	C	13	AAGGTAGG [AAGG]11	13	AAGGTAGG **[AAGG]11**
12 D2S441	F	11	[TCTA]11	16	[TCTA]13 TTTA[TCTA]2
	M	12	[TCTA]12 rs74640515 (G/A)	12	**[TCTA]12** rs74640515 (G/A)
	C	11	[TCTA]11	11	**[TCTA]11**
13 D5S818	F	12	[AGAT]12 rs25768 (A/G)	12	**[AGAT]12** rs25768 (A/G)
	M	11	[AGAT]11 rs25768 (A/G)	12	[AGAT]12 rs25768 (A/G)
	C	11	[AGAT]11 rs25768 (A/G)	13	**[AGAT]13** rs25768 (A/G)
14 D5S818	F	12	[AGAT]12 rs25768 (A/G)	12	**[AGAT]12** rs25768 (A/G)
	M	12	[AGAT]12 rs25768 (A/G)	12	[AGAT]12 rs25768 (A/G)
	C	12	[AGAT]12 rs25768 (A/G)	13	**[AGAT]13** rs25768 (A/G)
15 CSF1PO	F	10	[AGAT]10	11	[AGAT]11
	M	12	[AGAT]12	12	[AGAT]12
	C	8	[AGAT]8	12	[AGAT]12

Discussion

The NGS technology allowed to obtain a concordant result with those by CE system, showing the reliability of both methods. This methodology combined with the use of the Precision ID GlobalFiler™ NGS STR Panel v2 demonstrated to be useful to clarify the single exclusion, since it was possible to confirm the origin of the mutation by interpreting the sequence of the alleles for the majority of paternity trios.

Regardless of the trio, it was possible to verify that for mutations in the same genetic marker (FGA, vWA, D12S391 and D5S818), the repeat motif mutated was always the same, associated with the presence of an insertion or deletion.

For the D5S818 marker, it was observed the presence of the same SNP in the flanking region in both trios.

Finally, for the trio case with the silent or null allele, the discrepancy in allele inheritance may be due to a primer binding site mutation rather than a repeat mutation, since it was not detected in both parents or in any of the methodologies.

Conclusion

With the NGS methodology, it was possible to evaluate what type of mutation had occurred in the allelic transmission by analyzing the sequence data of the alleles, with the most frequent mutation being the insertion or deletion of a specific repeat unit.

Also, NGS provides additional information about the STR repeat motif sequence, which could increase the discrimination power of several genetic markers. The presence of 9 additional multi-allelic STRs in the NGS panel compared to the traditional CE method may be an important tool in paternity cases with more than one allelic transmission mismatch. In a future research, it will be important to include the allele frequencies of these additional markers to calculate the paternity index, in order to improve the statistical values. Therefore, the Precision ID GlobalFiler™ NGS STR Panel v2 shows to be a powerful method for kinship analyses and typing reference samples through a sequence-based analysis.

Conflict of interest statement

The authors have no conflict of interest to declare.

References

[1] H. Li, et al., Applying massively parallel sequencing to paternity testing on the Ion Torrent Personal Genome Machine, Forensic Science International: Genetics 31 (2017) 155-159.

[2] J. Wu, et al., Forensic application of a novel MPS-based panel (90 STRs and 100 SNPs) in a non-exclusion parentage case with three autosomal STRs incompatibilities, Legal Medicine 54 (2022) 101-987.

[3] A. Chen, et al., Investigation on the genetic-inconsistent paternity cases using the MiSeq FGx system, Forensic Sciences Research 7(4) (2022) 702–707.

[4] R. Fonseca, et al., Analysis of isoalleles and flanking SNPs of STR markers by NGS to distinguish monozygotic twins. Forensic Science International: Genetics Supplement Series 8 (2022) 285-287.

[5] L. Davenport, et al., Forensic identity SNPs: Characterisation of flanking region variation using massively parallel sequencing. Forensic Science International: Genetics 64 (2023) 102847.

[6] C. Faccinetto, et al., Internal validation study of the next generation sequencing of GlobalFiler™ PCR amplification kit for the Ion Torrent S5 sequencer, Forensic Science International: Genetics Supplement Series 7 (2019) 336-338.

[7] Thermo Fisher Scientific (2023). Precision ID GlobalFiler™ NGS STR Panel v2 with the HID Ion S5™/HID Ion GeneStudio™ S5 System Application Guide.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 747-753
DOI: https://dx.doi.org/10.15304/cc.2025.1869

STR allele dropout and flanking SNP analysis with The Precision ID GlobalFiler™ NGS STR Panel v2 kit in monozygotic twins

Cintia Fridman[1], Renata Inzinna Bernardo Fonseca[1],
Guilherme Valle-Silva[2], Celso Teixeira Mendes-Junior[2]

[1]Departamento de Medicina Legal, Bioética, Medicina do Trabalho e Medicina Física e Reabilitação, Faculdade de Medicina FMUSP, Universidade de São Paulo, São Paulo, SP, BR.

[2]Laboratório de Pesquisas Forenses e Genômicas, Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, BR

**Corresponding author: cfridman@usp.br*

Abstract

NGS (Next Generation Sequencing) is being validated for human identification to address limitations of Capillary Electrophoresis (CE), allowing for a detailed analysis of STR sequences, isoalleles, and flanking regions. However, no reports exist on using STR genotyping with NGS to differentiate monozygotic (MZ) twins, who are indistinguishable via CE. This study aimed to evaluate the potential of distinguishing MZ twins using the Precision ID GlobalFiler™ NGS STR Panel v2 by analyzing STRs and flanking regions. CE and NGS profiles of 32 MZ pairs were compared, and NGS results analyzed with different software. Our data showed that allele lengths were consistent between CE and NGS, except for one case with allele dropout in NGS. Isoalleles were identified in eight markers but did not differentiate twins. Flanking region analysis suggested unbalanced SNPs in two pairs, but further validation revealed these were sequencing errors. While unable to differentiate MZ twins, the NGS kit effectively identified isoalleles, enhancing discrimination power. Routine analysis with multiple software is recommended to avoid errors and confirm results.

Keywords

NGS, STRs, flanking SNPs, monozygotic twins

1. Introduction

While STRs (Short Tandem Repeats) are the gold standard in human identification, NGS (Next Generation Sequencing) technologies offer advantages by analyzing thousands of markers in a single reaction and these strategies are currently under validation in the field of human identification for forensic purposes, aiming to overcome some limitations of Capillary Electrophoresis (CE), allowing a comprehensive evaluation of STRs' sequences, their isoalleles and flanking regions [1–4].

Despite these advancements, there is limited data on their effectiveness in differentiating monozygotic (MZ) twins, who arise from a single fertilized egg. In these cases, minor genetic differences may emerge depending on the timing of embryo division [5–7].

Since the birth rate of twins has increased worldwide due to factors such as the use of in vitro fertilization techniques and advanced maternal ages, improvements in this field are particularly important [8,9]. In Brazil, from 2005 to 2023, approximately 80 legal cases involving identical twins (both suspects and victims) were reported. This study aimed to evaluated the possibility to distinguish MZ using the Precision ID GlobalFiler™ NGS STR Panel v2 by analyzing STRs and their flanking regions.

2. Material and Methods

Peripheral blood was collected from 32 healthy MZ pairs, aged 18 to 65. DNA extraction was performed using the Salting-out technique, and quantification with the Nanodrop 2000 (ThermoFisher). All volunteers signed an informed consent, and the research protocol was approved by the Ethics Committee of the Clinical Hospital, University of São Paulo (CAEE 81950617.7.0000.0065).

Zygosity was confirmed by STR profiling using the Power Plex® Fusion System kit (Promega). Amplified products were analyzed via CE on the ABI 3130 (Applied Biosystems) with genotypes processed in GeneMapper ID v3.2.

For NGS, library construction, template preparation, sequencing, and data analysis were performed using the Precision ID GlobalFiler™ NGS STR Panel v2 kit (ThermoFisher) on Ion Chef and Ion S5 Gene Studio equipment. Raw NGS data were analyzed with Converge™ v2.1 software and validated using Integrative Genomics Viewer (IGV) v2.12.3. In a secondary analysis, FastQ files were processed with STRait Razor v3 software. Sequencing error rates were analyzed in Microsoft Excel (Office 365).

To confirm the SNP in TPOX and D6S1043 markers, specific primers were designed using Primer3 and sequencing reactions were conducted using BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) and analyzed via capillary electrophoresis on ABI3130, with results compared in BioEdit Sequence Alignment Editor v7.7.1 to NGS data.

3. Results

The CE and NGS profiles were compared for all volunteers and the allele results were concordant, except for one dropout allele on Penta D sequencing data analyzed with Converge software, in which CE results of G037 pair showed allele 10 and 13 for Penta D marker, but Converge only showed allele 10 for G037B individual. Despite NGS results, IGV and STRait Razor software were able to detect both alleles, 10 and 13, even with low coverage (19 reads in allele 13 on STRait Razor results).

Isoalleles were observed in ten pairs of MZ twins, with genotypes that were initially described as homozygous by CE: G021 (D2S441 marker), G040 (D12S391 marker), G041 (D4S2408 marker), G054 (D12S391 and D3S1358 markers), G065 (D3S1358 and D8S1179 markers), G068 (D3S1358 marker), G069 (D21S11 and D3S1358 markers), G076 (D3S1358 marker), G087 (D3S1358 marker), and G097 (D12ATA63 and vWA). Those isometric alleles were also found by other published papers with different populations, as Brazilian, Indian, Spanish, and Chinese [4,10–14].

Finally, we have previously observed and published [15] data regarding Converge results analysis showing two flanking unbalanced heterozygous SNP genotypes in these sample, each one in a pair, in only one individual of the pair: rs560609904 on allele 8 in TPOX (G016A) and rs569521603 on allele 13 in D6S1043 (G027B). However, to further comprehend this finding, we extended the analysis by examining these positions in detail using IGV. This new evaluation revealed that only 6.09% of the sequences displayed the G>A substitution at rs560609904 (TPOX), and only 2.43% of the reads exhibited the G>T substitution at rs569521603 (D6S1043) in individuals G016A and G027B, respectively. Next, we also analyzed these possible SNP with STRait Razor, in which for TPOX, the coverage of reads containing the Guanine base was 1073, while those with alteration to Adenine were only 77, representing only 7.17% of the total number of reads. For D6S1043, the coverage of the sequences with Guanine was 4149, while the coverage of the ones with Timine was only 102 reads, representing 2,45% of the total number of reds.

To determine whether these observations detected by the three software tools (Converge, IGV and STRait Razor) were the result of sequencing errors rather than true somatic events, Sanger sequencing was performed to verify the findings. By the analysis of forward strand in TPOX and reverse strand in D6S1043, no signs of alternative bases at the investigated positions were detected, indicating that the initial observations by Converge were likely due to sequencing errors rather than somatic mutations.

Since the SNPs pointed out by Converge were not confirmed, we evaluated the error rates in NGS for these two markers. Regarding TPOX, allele 8 on individual G016A presented an error rate of 31.70%, while the error rate for this allele on G016B was 7.69%. The D6S1043 allele 13 sequencing error rate was 13.78% and 11.36% for individuals G027B and G027A, respectively. Taken together, the elevated sequencing error rates highlight the high probability of sequencing errors being misinterpreted as somatic mutations.

4. Discussion

The comparison of the CE and NGS results of the 32 pairs of MZ twins showed concordance between them, except for one allele dropout on Penta D marker, which occurred probably because its low coverage on NGS, affecting the experiment quality. Studies have shown low-performance values parameters of allele coverage for Penta D marker, highlighting the need of improvement of these markers for future versions of the kits [4,16].

We also found 22 isoalleles, in 8 different markers, that did not differentiate monozygotic twins. This data corroborates NGS as an excellent technique for detecting variations found in the internal sequences of isometric alleles. Other studies have found the same isoalleles as ours [4,10–13,16], but there is no consensus regarding the nomenclature of these variations, so we believe that it is necessary to create and standardize a database for existing variations, making their use feasible. The presence of isoalleles in our sample highlight the possibility of using them to increase the power of discrimination in population analyses.

Two flanking SNPs were observed in NGS results in only one of the twins in each of the pairs in our study [15], but after additional analysis with other software and Sanger sequencing, these data were not confirmed, suggesting that the alternative bases observed was due to sequencing errors.

5. Conclusion

Despite being unsuccessful in differentiating monozygotic twins in this current sample, we have observed that the Precision ID GlobalFilerTM NGS STR Panel v2 kit is indeed effective at identifying isoalleles, which increases the combined discrimination power. Notwithstanding, we strongly recommend routine analysis of NGS data with more than one genotype calling software in order to avoid allele dropout events and to confirm the results, taking extreme care to tell apart sequencing errors from actual genetic variability.

6. Acknowledgments

This study was supported by FAPESP (20/08092-5), CAPES and HCFMUSP/LIM40. We would like to thank ThermoFisher Scientific for technical support.

7. Conflict of interest statement

None

8. References

[1] Butler JM. Advanced Topics in Forensic DNA Typing: Methodology. 1st ed. Maryland: Academic Press; 2011. https://doi.org/10.1016/B978-0-12-374513-2.00003-8.

[2] Gettings KB, Kiesler KM, Faith SA, Montano E, Baker CH, Young BA, et al. Sequence variation of 22 autosomal STR loci detected by next generation sequencing. Forensic Sci Int Genet 2016;21:15–21. https://doi.org/10.1016/j.fsigen.2015.11.005.

[3] Parson W, Ballard D, Budowle B, Butler JM, Gettings KB, Gill P, et al. Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. Forensic Sci Int Genet 2016;22:54–63. https://doi.org/10.1016/j.fsigen.2016.01.009.

[4] Tao R, Qi W, Chen C, Zhang J, Yang Z, Song W, et al. Pilot study for forensic evaluations of the Precision ID GlobalFiler™ NGS STR Panel v2 with the Ion S5™ system. Forensic Sci Int Genet 2019;43:8. https://doi.org/10.1016/j.fsigen.2019.102147.

[5] Krawczak M, Budowle B, Weber-Lehmann J, Rolf B. Distinguishing genetically between the germlines of male monozygotic twins. PLoS Genet 2018;14:1–9. https://doi.org/10.1371/journal.pgen.1007756.

[6] McNamara HC, Kane SC, Craig JM, Short R V., Umstad MP. A review of the mechanisms and evidence for typical and atypical twinning. Am J Obstet Gynecol 2016;214:172–91. https://doi.org/10.1016/j.ajog.2015.10.930.

[7] Yuan L, Chen X, Liu Z, Liu Q, Song A, Bao G, et al. Identification of the perpetrator among identical twins using next-generation sequencing technology: A case report. Forensic Sci Int Genet 2020;44:6. https://doi.org/10.1016/j.fsigen.2019.102167.

[8] Pison G, Monden C, Smits J. Twinning Rates in Developed Countries: Trends and Explanations. Popul Dev Rev 2015;41:629–49. https://doi.org/10.1111/j.1728-4457.2015.00088.x.

[9] Torres C, Caporali A, Pison G. The Human Multiple Births Database (HMBD). Demogr Res 2023;48:89–106. https://doi.org/10.4054/demres.2023.48.4.

[10] Silva DSBS, Sawitzki FR, Scheible MKR, Bailey SF, Alho CS, Faith SA. Genetic analysis of Southern Brazil subjects using the PowerSeq™ AUTO/Y system for short tandem repeat sequencing. Forensic Sci Int Genet 2018;33:129–35. https://doi.org/10.1016/j.fsigen.2017.12.008.

[11] Barrio PA, Martín P, Alonso A, Müller P, Bodner M, Berger B, et al. Massively parallel sequence data of 31 autosomal STR loci from 496 Spanish individuals revealed concordance with CE-STR technology and enhanced discrimination power. Forensic Sci Int Genet 2019;42:49–55. https://doi.org/10.1016/j.fsigen.2019.06.009.

[12] Moura-Neto R, King JL, Mello I, Dias V, Crysup B, Woerner AE, et al. Evaluation of Promega PowerSeq™ Auto/Y systems prototype on an admixed sample of Rio de Janeiro, Brazil: Population data, sensitivity, stutter and mixture studies. Forensic Sci Int Genet 2021;53:8. https://doi.org/10.1016/j.fsigen.2021.102516.

[13] Kwon YL, Kim BM, Lee EY, Shin KJ. Massively parallel sequencing of 25 autosomal STRs including SE33 in four population groups for forensic applications. Sci Rep 2021;11:9. https://doi.org/10.1038/s41598-021-82814-z.

[14] Dash HR, Kaitholia K, Kumawat RK, Singh AK, Shrivastava P, Chaubey G, et al. Sequence variations, flanking region mutations, and allele frequency at 31

autosomal STRs in the central Indian population by next generation sequencing (NGS). Sci Rep 2021;11. https://doi.org/10.1038/s41598-021-02690-5.

[15] Fonseca RIB, Fridman C. Analysis of isoalleles and flanking SNPs of STR markers by NGS to distinguish monozygotic twins. Forensic Sci Int Genet Suppl Ser 2022. https://doi.org/10.1016/j.fsigss.2022.10.064.

[16] Dash HR, Rawat N, Vajpayee K, Shrivastava P, Das S. Useful autosomal STR marker sets for forensic and paternity applications in the Central Indian population. Ann Hum Biol 2021;48:37–48. https://doi.org/10.1080/03014460.2021.1877353.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 755-761
DOI: https://dx.doi.org/10.15304/cc.2025.1869

On the use of forensic investigative genetic genealogy in criminal cases in Sweden

Ricky Ansell[1,2*], Siri Aili Fagerholm[1]

[1]*The Swedish Police Authority, National Forensic Centre, Linköping, Sweden*
[2]*Department of Physics, Chemistry & Biology (IFM), Linköping University, Linköping, Sweden*

** Corresponding author at: Swedish Police Authority, National Forensic Centre, SE-581 94 Linköping, SWEDEN, ricky.ansell@polisen.se*

Abstract
Sweden was in 2019, the first country outside of North America in solving a crime case using forensic investigative genetic genealogy (FIGG), but further use of the method was inhibited by the Privacy Protection Authority (IMY) in 2021 during a mandatory pre-counselling process, due to considered lack of legislative support and issues regarding transfer of sensitive data to a third country. Subsequently, method implementation was set on halt awaiting necessary legal review and amendments. In July 2024, the Swedish Justice Department presented a legal bill on the police use of FIGG, following the outcome of a national biometric committee presenting their assessments in June 2023. This paper describes the proposed legislation on the use of FIGG, presenting the proposed legal amendments and discussing some underlying considerations.

Keywords
FIGG, Forensic Investigative Genetic Genealogy, DNA, Legislation.

Introduction
Sweden was, outside of North America, first in using forensic investigative genetic genealogy (FIGG), to try identify human remains in a 2003 murder case in late 2018 [1], and early 2019 in a formalised pilot case to obtain leads on the culprit in an unsolved 2004 double murder case [2-4]. The double murder case was successfully solved in midst of 2020, while the identification of the victim was not successful and the case remains unsolved.

In 2021, following a mandatory pre-counselling process, the ongoing implementation of the method within Swedish police, establishing practical routines and national guidelines on the use of FIGG, was inhibited by the supervisory authority IMY (Integritetskyddsmyndigheten: Swedish Authority for Privacy Protection) [4]. Main issues addressed by IMY were lack of legal support for third-party database searches and processing of sensitive personal data, as well as issues around the transfer of sensitive data to a "third country".

In parallel, a national committee on modernizing and harmonizing legislation for law enforcement use of different biometrics modalities had, for completely other reasons, been launched. Early 2022, the Justice Department added to the committee the task to perform an overview on the use of FIGG, to consider and present legislative changes required to allow for its use by the Swedish Police Authority. Notable is that the task did not cover the use of FIGG in identifying unidentified victims of crime or human remains but only the use of crime scene DNA to identify a perpetrator. Repeated attempts to include identification cases were not successful.

The committee presented their outcome in June 2023 [5]. Subsequently, Justice Department processed a legal bill presented in July 2024 [6], proposing a legislation on the use of FIGG. The bill is expected to pass a parliament vote, but has to pass through the Council of Legislation (Lagrådet) for review prior to finalizing the bill and passing a vote. The legal bill mirrors in most parts the national committee report, which in turn adopted most conclusions drawn from the pilot case report [3,5,6].

In the following the proposed legislation and the essence of the legal bill - shaping the landscape for the use of FIGG in Sweden - is presented and discussed.

Material

The material used for this work is the legal bill with paragraphs and constitutional comments: "Lagrådsremiss - Biometri i brottsbekämpningen" [6].

Results

Proposed legislation

Excerpt of the legal bill proposal for amending present legislation "*Law (2018:1693) on the police's processing of personal data in the area of the Criminal Data Act*":

6 Ch. Processing of personal data at the Police Authority for forensic purposes

§1 At the National Forensic Centre, personal data may be processed if it is necessary to

… …

7. use a DNA-based genealogy database to perform comparisons according to Ch. 6b.

6b Ch. Processing of data in DNA-based genealogy databases

§1 The National Forensic Centre may, after a decision by the prosecutor, use a DNA-based genealogy database to compare information about DNA from traces with information in the database.

§2 A decision to use a DNA-based genealogy database can be made during a preliminary investigation in murder according to Ch. 3. Section 1 of the Criminal Code and aggravated rape or aggravated rape against a child according to Ch. 6 Section 1, third paragraph or Section 4, third paragraph of the Criminal Code if

1. there is special reason to assume that the DNA trace comes from the perpetrator,
2. the measure is of particular importance to the investigation and it is clear that its purpose cannot be achieved with less invasive measures.

§3 In order for a DNA-based genealogy database to be used, the Police Authority must ensure that

1. the provider of the genealogy database does not process the data for other purposes,
2. the comparison only includes data on DNA from persons who have agreed that their data is processed to investigate crimes,
3. the provider of the genealogy database deletes the data when the Police Authority requests it.

§4 The search ban in Ch. 2 Section 14 of the Criminal Data Act (2018:1177) does not prevent searches according to this chapter for the purpose of obtaining a selection of persons based on genetic or biometric data.

The requirements set out in the third paragraph apply in addition to what otherwise follows from the Data Protection Regulation, including the requirements in Ch. 3. the Criminal Data Act (2018:1177) on the obligations of the person responsible for handling of personal data and Ch. 8 (the same law) on the transfer of personal data to a third country.

Table 1. Timeline with the major steps towards implementing FIGG in Sweden. Activities in italics are ongoing or coming. (IMY - Integritetsskyddsmyndigheten: Swedish Authority for Privacy Protection).

Activity	
Legal inquiry	2018-2019
Method set-up	2018-2019
Police Ethics Board	2018
Pilot case (Crime scene sample)	2019-2020
Pilot case court verdict [a]	2020
Pilot evaluation	2020-2021
Pre-counselling at IMY	2021
Biometric committee	2022-2023
Draft legal bill (Lagrådsremiss)	2023-2024
Council on Legislation	2024
Legal bill (Proposition)	*2024*
Parliament vote	*2024*
Pre-counselling at IMY	*2025*
Legislation in force	*2025*
Implementation	*2020 -*

[a] The verdict was not appealed.

Discussion

The proposed legislation provides provision lifting the search ban on the police use of third party DNA-based genealogy databases for some of the most severe crimes, in order to obtain a selection of individuals based on genetic data to be used for further processing. The suggested legislation and underlying proceedings ("constitutional comments") are written with clear limitations aiming at restrictiveness in its use.

Due to its sensitive nature the lawmakers have shaped paragraphs that are specific for the use of FIGG. Using FIGG is argued as being exceptional and presupposes both "absolute necessity" and that other investigative measures have been

exhausted without success, including extensive use of DNA (familial searches[3], phenotyping & ancestry etc.).

Transfer to a private actor in third country (and not to a mandated authority) is doable but only if fulfilling requirements set in the Criminal Data Act. Also, a decision on using FIGG is to be taken on a case-by-case basis. In the legal bill it is stated that "Investigative genetic genealogy constitutes forensic activity, which NFC (National Forensic Centre) is tasked with carrying out". The background is that within the Police Authority genetic data is only allowed to be processed by the NFC. As genealogy will make use of DNA data, the work performed by a genealogist is thus concluded to be performed under the umbrella of NFC. If genealogists are to be employed, contracted or both remains to be seen.

One major difference appearing in the legal bill, compared with the pilot study, regards the involvement of a prosecutor to decide on the use of the method. Many of the cases in question will be police-led (as there is no suspect at the time), which means that a prosecutor now must be enrolled, at least temporarily in deciding the use of FIGG. This is a clear statement from the lawmakers on restrictiveness: Some consultation bodies reviewing the committee report suggested an even stricter process involving a court decision for each specific case.

Furthermore, the prosecutor will need to seek advice "high within the Police Authority" prior to a decision on using FIGG. This to assure that all criteria and obligations set on the police are met. Also, prioritization in between cases will be handled within the police. At this stage it is still unclear how this advisory function will be formed, but it is obvious that different expertise's will need to be involved, apart from prosecutors and highly ranked police officers.

Even with a legal framework present, national policy documents on the use of FIGG will need to be developed. Of special relevance is describing the interface at which the prosecutor enters the stage and the required consultation with the police, prior to a decision on using FIGG.

One major limitation and drawback with the proposed legislation is that the provisions do not allow a third-party DNA-based database to be used for the purpose of identifying a deceased victim or human remains.

The proposed legislation is suggested to come active July 1st 2025.

---

[3] The initial aim with the pilot case was in fact to handle two unsolved cases using FIGG to find the perpetrator, though only one crime case ended up being used. The second crime was a severe rape case against a child in 1995, and this case was solved early 2019, following new legislation allowing familial searches in the police DNA databases – thereby adding arguments on using existing "DNA tools" before the use of FIGG.

Conclusion

As outlined in Table 1, the road towards legislation is long and takes considerable time. In Sweden, the only way forward was by introducing new legislation as the police use of third-party databases was not covered by present legislation. The proposed legislation is restrictive mirroring the view that the use of FIGG is regarded as sensitive and considered only to be a tool of last resort. In most parts, statutes as well as constitutional comments and considerations follow the suggestions and arguments described already in the pilot case report and later by the Biometric inquiry. Our view is that a transparent pilot study did pave the way and have been subsequently used in forming the proposed legislation.

Conflict of interest statement

The authors declare no conflict of interest.

References

[1] A. Tillmar, P. Sjölund, B. Lundqvist et al. Whole-genome sequencing of human remains to enable genealogy DNA database searches – A case report. Forensic Sci. Int. Genet. 46 (2020) 102233, https://doi.org/10.1016/j.fsigen.2020.102233

[2] A. Tillmar, S. Aili Fagerholm, J. Staaf et al. Getting the conclusive lead with investigative genetic genealogy – A successful case study of a 16 year old double murder in Sweden. Forensic Sci. Int. Genet. 53 (2021) 102525, https://doi.org/10.1016/j.fsigen.2021.102525

[3] S. Fagerholm, R. Ansell, A. Tillmar et al. Pilot: Dna-spår och släktforskning. Användning av släktforskningsdatabaser i brottsutredande syfte. Polismyndigheten, NFC, November 2020, A544/825/2020, https://polisen.se/siteassets/dokument/forensik/rapport-pilotprojekt-slaktforskningsdatabaser-v1.0-1.pdf (Accessed 2024-09-23)

[4] S. Fagerholm, A. Tillmar, J. Staaf et al. Forensic DNA traces and genealogy, Use of investigative genetic genealogy in criminal investigations. Polismyndigheten, NFC, November 2021, A594/805/2021, https://polisen.se/siteassets/dokument/forensik/forensic-dna-traces-and-genealogy.pdf (Accessed 2024-09-23)

[5] Biometri – för en effektivare brottsbekämpning, SOU 2023:32. Regeringskansliet, 19 juni 2023, https://www.regeringen.se/contentassets/08af26f9ffca495b8bd9c2f0c77e169d/biometri--for-en-effektivare-brottsbekampning-sou-2023-32.pdf (Accessed 2024-09-23)

[6] *Lagrådsremiss - Biometri i brottsbekämpningen. Regeringskansliet, 4 juli 2024,* https://www.regeringen.se/rattsliga-dokument/lagradsremiss/2024/07/biometri-i-brottsbekampningen/ (Accessed 2024-09-23)

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 763-768
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Experiences of the use of combined DNA and mRNA analysis in criminal cases

Siri Aili Fagerholm[1*], Ricky Ansell[1,2], Christina Forsberg[1]

[1] *National Forensic Centre, Swedish Police Authority, Linköping, Sweden.*

[2] *Department of Physics, Chemistry & Biology, Linköping University, Linköping, Sweden.*

** Corresponding author at: National Forensic Centre, Swedish Police Authority, SE-581 94 Linköping, SWEDEN, siri.aili-fagerholm@polisen.se.*

Abstract

In forensic criminal investigations, a DNA result may not be sufficiently informative unless there is additional information about the body tissue or cell type/body fluid from which the extracted DNA originates. Especially in criminal cases, in which the suspect and victim has had social or physical contact before the assault, it can be of vital importance for the case to investigate the nature or possible source of the DNA findings. One method that may be used is DNA analysis followed by cell type specific mRNA analysis. For the past years the Netherlands Forensic Institute (NFI) has performed DNA/mRNA analysis on request by the Swedish National Forensic Centre (NFC). In total, samples from more than 60 sexual assault cases were analyzed between 2015 to 2023. In the vast majority of cases, swabs from the suspects fingers/nail scrapings were analyzed in order to detect possible presence of DNA from the victim. If relevant DNA was detected a subset of the samples was further analyzed with mRNA typing for the possible presence of predominantly vaginal cells and/or menstrual secretion. In approximately half of the criminal cases, DNA of interest was detected and samples subjected to RNA typing analysis. Approximately in half of these cases, an indication of the presence of RNA for a specific cell type were found. For a number of cases the findings were of significant value for the prosecutors claims. In conclusion, combined DNA/mRNA analysis has been successfully performed on samples from predominately sexual crime cases in which social and/or physical contact preceded the assault. When DNA/mRNA findings were later presented as evidence in Swedish courts, they were regarded as moderate to strong evidence.

Keywords

mRNA, body fluid identification, BFI, sexual assaults.

Introduction

In recent years the possibility to perform identification of cell types/body fluids or organ tissues using mRNA typing has emerged as a useful tool in forensic criminal investigations. Using cell type specific RNA markers it may be possible to link extracted DNA to a certain cell type such as vaginal secretion or saliva. In criminal cases in which the suspect and victim has had social or physical contact prior to an assault it can be of vital importance for the case to investigate the nature or possible source of the DNA findings as DNA may have transferred during the legal contact only and not during the assault.

For the past years the Netherlands Forensic Institute (NFI) has performed DNA analysis followed by mRNA analysis on request by the Swedish National Forensic Centre (NFC). In this paper, the outcome and experienced gained will be discussed.

Material studied

Case reports from the NFI for all criminal cases in which material has been subjected to DNA/mRNA examination during the years 2015-2023 were examined. Case verdicts from the District Courts and the Courts of Appeal, has been accessed from the online database Juno.

Results

During a period of nine years (2015-2023) requested analyses has been performed at the NFI in 65 cases and, when relevant DNA was detected a subset (if applicable) of samples were subjected to mRNA typing using NFI's methods [1]. With mRNA typing the presence of vaginal cells, menstrual secretion, saliva, blood, semen and nasal mucosa has been analyzed. Positive findings were reported as an indication of presence of a certain cell type. During the last years (since 2021), the weight of evidence of the mRNA findings of vaginal/menstrual secretion has been calculated and presented in verbal terms in the case reports. In one case, DNA analysis was instead combined with mRNA organ tissue typing [2,3] and the possible presence of adipose tissue, central nervous system, lung, liver, skeletal muscle, heart and kidney tissue were examined.

Some inclusion criteria were used before samples were submitted for DNA/mRNA analysis and these included that the crime had to be a serious offence, such as sexual assault or murder. All cases but one were classified as sexual assaults and the possible presence of vaginal/menstrual secretion, saliva or semen was in question. The exception was a murder case (gun violence) in which DNA/organ tissue typing (as mentioned above) was performed on a swab with biological material from the crime scene, in an attempt to position the shooting incident to a specific location.

Trace recovery was recommended to have been performed within 24 hours of the alleged crime for swabs from a person's body, such as fingernails or penile swabs. In some cases, samples taken after 24 hours were still submitted on request by the prosecutor and in a few of these samples relevant DNA was detected, though none produced any relevant RNA result.

In each case the investigation team and prosecutor were informed of the cost of the investigation (as it is substantially higher than regular DNA analysis), expected delivery time (longer than for regular DNA analysis) as well as possible outcomes and reporting procedures. The investigation leader thereafter had to make an informed decision of submitting material for DNA/mRNA analysis.

Samples ranged from finger/nail scrapings (68% of total numbers of examinations), underwear, lip and penile swabs from suspect (11%), underwear and vaginal swabs from victim (6%), different objects such as sex toys and other items used for vaginal penetration (6%), condoms (3%) and other items such as swabs from a steering wheel, stains on a pillow case and a napkin (6%). When possible, a pre-screening procedure of samples for relevant donor DNA was performed at NFC. This enabled us to exclude samples for further analysis that did not display DNA from the relevant person.

In approximately 50% (42/78) of the examinations relevant DNA, from the victim or suspect depending on case circumstances, were found (Table 1). Both relevant DNA and mRNA was detected in about 30% (26/78) of the examinations. Note: more than one item may have been analyzed per examination and case (e.g. five swabs from suspects right hand is counted as one examination). Also, a selection procedure was used by the NFI on samples with relevant DNA for which a subset was further analyzed for mRNA expression.

Positive likelihood ratios (LR) for mRNA originating from vaginal cells and/or menstrual secretion and possible other types of cell(s) (with the alternative proposition that a sample contained only other types of cell(s)) ranged between LR 1 – 10,000.

Table 1. Overview of samples analyzed and results. Sample type, total number of examinations (more than one item may have been analyzed per examination), number of examinations in which no relevant DNA was detected, relevant DNA and no relevant mRNA, or relevant DNA/mRNA results.

Sample type	Total number of examinations	No relevant DNA	Relevant DNA, no relevant mRNA	Relevant DNA and mRNA
Finger/nail scrapings (suspect)	53	25	12	16
Lip swabs (suspect)	3	3	-	-
Underwear, penile swabs (suspect)	5	1	1	3
Underwear, vaginal swabs (victim)	5	4	-	1
Objects	5	-	2	3
Condoms	2	1	-	1
Other	5	2	1	2
Total	78	36	16	26

Of investigative reasons, not all DNA or DNA/mRNA findings were later used and presented as evidence in court. In the court verdicts findings were summarized and discussed. In common, the court reasoned that the findings constitute one piece of evidence and that it provides a support to other evidence such as the victim´s statement. In the verdicts it is reasoned that the findings range from providing "some support" to "strong support" for the prosecutors claims.

To address the findings presented in court, it is clear from the verdicts that the defendants provided different explanations to the results. The explanations mainly followed the outlined scenarios given by defendants in Dutch court cases [1]. First, the defense provided an innocent explanation to explain DNA/mRNA findings, e.g. suspect claimed to only have been making out with the victim. Second, the results where explained by possible secondary transfer of DNA, e.g. suspect claimed to only have massaged the woman's feet, and denied vaginal penetration with fingers. Third, the defense referred the results as a consequence of a time of deposition mechanism, e.g. suspect having previous sex with another woman or having sex at a previous time point with the same woman.

Investigators and prosecutors were asked about their opinion of the examinations and perceived value in court. As a new method for most, the perception was that it took a long time to grasp the method and findings, and discussions were often initiated between the investigation team and reporting officers at NFC and/

or NFI. Though in general, prosecutors and investigators responded that they had found the conclusions to have been of great importance in the investigation of the alleged crime and in the court proceedings.

Discussion

It is obvious that DNA in combination with mRNA/cell type results can provide added value to an investigation compared to a DNA result only. Especially in sexual assault cases in which the victim and suspect spent time together before the assault and in which DNA results may be explained by DNA transfer during the legal undisputed contact. For these cases, the benefit of performing a DNA/mRNA analysis is well motivated even when considering the higher cost and increased turnaround time.

In the future, possibilities to link mRNA results to a specific individual will add value and increase usefulness of the method. At present the method cannot help to distinguish between scenarios such as from whom of several women mRNA/vaginal secretion may be derived, or from which donor in a DNA mixture detected saliva may be derived. Potentially, coding region SNP-assays (cSNPs) for cell types/body fluids can be used to link a cell type to a specific donor [4].

With the focus on sexual assault cases, it would be desirable to increase the panel for mRNA typing to include markers also for e.g. rectal mucosa. To detect rectal mucosa would be of importance in sexual assaults with alleged anal penetration. Although markers for rectal mRNA have been explored [5-7], issues with cross-reactivity have been detected [6,7].

Conclusion

Combined DNA/mRNA analysis has been successfully performed in a number of samples from predominately sexual assault cases in which social and/or physical contact preceded the assault. When DNA/mRNA findings were later presented as evidence in Swedish courts, the findings were found to have high evidentiary value.

Conflict of interest statement

The authors declare no conflict of interest.

References

[1] T. Sijen and S. Harbison. On the identification of body fluids and tissues: A crucial link in the investigation and solution of crime. Genes (Basel) 12 (2021) 1728.

[2] A. Lindenbergh, M. van den Berge, R-J. Oostra et al. Development of a mRNA profiling multiplex for the inference of organ tissues. Int. J. Legal Med. 127 (2013) 891-900.

[3] M. van den Berge, and T. Sijen. Addition of an adipose marker to the organ tissue identification mRNA assay. Forensic Sci. Int. Genet. Sup. Ser. 8 (2022) 32-34.

[4] E. Hanson, G. Dørum, M. Zamborlin et al. Targeted S5 RNA sequencing assay for the identification of and direct association of common body fluids with DNA donors in mixtures. Int. J. Legal Med. 137 (2023) 13-32.

[5] T. Akutsu, H. Saito, K. Watanabe et al. Evaluation and simultaneous determination of rectal mucosa markers by multiplex reverse transcription-PCR for biological evidence of sexual assault with anal penetration. Forensic Sci. Int. Genet. 59 (2022) 102712.

[6] M. Neis, T. Groß, H. Schneider et al. Comprehensive body fluid identification and contributor assignment by combining targeted sequencing of mRNA and coding region SNPs. Forensic Sci. Int Genet. 73 (2024) 103125.

[7] M. Bamberg, L. Dierig, G. Kulstein et al. Development and validation of an mRNA-based multiplex body fluid identification workflow and a rectal mucosa marker pilot study. Forensic Sci. Int Genet. 54 (2021) 102542.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 769-773
DOI: https://dx.doi.org/10.15304/cc.2025.1869

A decade of mRNA body fluid and organ typing in casework: outcomes of Dutch court decisions

Titia Sijen[1*] and Margreet van den Berge[1]

[1]*Netherlands Forensic Institute, The Hague, The Netherlands*

**Corresponding author at: Laan van Ypenburg, 2497GB, The Hague, The Netherlands, t.sijen@nfi.nl*

Abstract
RNA typing to infer which body fluids or organs reside in an evidentiary stain, has been applied routinely in casework at our institute since 2012. Here, we regard verdicts in 33 cases that considered RNA results in the trials as described in the Dutch registry of verdicts [1]. We focused on the core of the alternative scenario put forward by the defense as this may feed our research.

Keywords
Forensic, mRNA, body fluid, organ, court decision

Introduction

The process of RNA typing starts with co-extracting DNA and RNA from the same sample: when the DNA is bound to a silica-containing column, the RNA goes to the flow-through and can be extracted separately. The RNA molecules remain stable for years when RNA extracts (or flow-through fractions) are stored at -80°C. This allows for the selection of the most informative samplings based on the DNA-profiling results or for securing RNA when it is not yet decided that RNA typing is useful in a case. RNA typing may relate either to body fluid or organ type inference, which are two different assays in our laboratory. Since 2012, we have processed 3252 samples in 921 cases; 61.6% of the RNA samples were stored, 26.6% were taken to body fluid typing and 11.9% to organ typing (for cases the percentages are: storage in 50.1%, body fluid typing in 36.5%, organ typing in 12.5% and both organ and body fluid typing in 1.0% of the cases).

Methodology

In the Netherlands, around 40% of the ~26,000 criminal cases that are yearly handled by courts are published (anonymously) online in the Dutch registry of verdicts. These court decisions were searched with the term 'RNA' and the contents were screened for the role of the RNA results in the ruling as a follow-up of [2]. With appeals in a case, the most recent ruling was taken.

Results

Thirty-three court decisions were found, as summarized in Table 1. In 27 cases (numbers 1-27), the RNA result was used in the decision by the judge; in six cases (numbers 28-33) not, which could either be because there was no RNA result (cases 31-33) or because the RNA typing was not informative for the question central in the court case (see cases 28-30), even though the RNA typing results may have been useful in formulating hypotheses in the police investigations. The defense raises various alternative scenarios that can be grouped into four main categories: 1) time of deposition (three cases); 2) secondary transfer (seven cases, all sexual assaults); 3) an innocent (or less invasive) explanation (11 cases) and 4) another person performed the crime (10 cases, all violent crimes). No alternative was presented in two cases. Two cases involved acquittal: the sexual abuse component in case 10, as it was unclear who is the donor of the saliva on the penis, and case 18 where it cannot be established that the suspect performed the stabbing although blood and muscle and a DNA match to the victim were found on the knife.

Discussion

For case 10, in which there was acquittal for the sexual abuse component, the association of donor and cell type would have been helpful. SNPs residing in transcribed regions (5'/3' UTR or coding region) of tissue-specific genes can facilitate this: the DNA will tell whether a donor (suspect or victim) carries the SNP; the RNA will inform whose body fluid (in case 10 saliva) was present in the evidentiary stain (the penile sampling) [5].

Table 1. Brief summary of court rulings found in Dutch registry of verdicts that mention RNA typing.

#	C: Case scenario A: Alternative scenario put forward by defense	RNA results[1, 2]	European Case Law Identifier	Verdict
1	C: Sexual assault on Friday, consensual intercourse on Monday before, period started on Tuesday A: Not clear **WHEN** cell material was deposited	Menstrual on penis	ECLI:NL:RBZLY: 2012:BW7487	Convicted
2	C: Homicide of elderly couple on boat A: Suspect visited boat and saw victims the **DAY AFTER** the crime	Blood, nasal on trousers suspect; pattern of expirated blood	ECLI:NL:GHARL: 2015:4875	Convicted
3	C: Murder husband by 70-yrs old wife, stabbed in heart, wounds hidden by clothing (wife phones 'he was stabbed'), towel nearby (used to clean knife?) A: Unclear **WHEN** the blood was deposited as a towel is a moveable object so relation crime unclear	Muscle, blood (match victim) on towel. Also, DNA matching suspect and fibers clothing victim on towel	ECLI:NL:GHARL: 2021:11501	Convicted
4	C: Sexual assault of unconscious granddaughter A: Assisted victim feeling unwell, then **TRANSFER** of her cell material to penis/underpants after urinating	Vaginal in underpants male suspect	ECLI:NL:GHARL: 2014:5505	Convicted
5	C: Sexual assault 16-yrs old daughter A: There are alternative ways like indirect **TRANSFER** how vaginal could get on fingers	Vaginal on fingers male suspect	ECLI:NL:RBNNE: 2015:2145	Convicted
6	C: Sexual assault by 2 men in bar: anal, vaginal, oral A: The suspect has touched object carrying vaginal cell material victim and **TRANSFERRED** this to penis	Vaginal, semen, saliva on penis suspect 1	ECLI:NL:RBROT: 2019:3840	Convicted
7	C: Sexual assault 13-yrs old daughter A: Suspect touched an area close to vagina victim and then **TRANSFERRED** vaginal cells to penis	Vaginal, semen, skin on penis suspect	ECLI:NL:RBAMS: 2020:3128	Convicted
8	C: Kidnapping from playground and sexual assault 5-yrs old girl, penetration of labia with penis A: **TRANSFER** vaginal to penis from touching (court: unlikely as no vaginal on hands suspect)	Vaginal on penis suspect (presumptive test indicates seminal fluid on labia victim)	ECLI:NL:RBOVE: 2021:1370	Convicted
9	C: Sexual assault 13-yrs old stepdaughter A: Vaginal on fingers, **TRANSFERRED** when suspect collected laundry of victim; underneath nails as he next moistened his hands, then to penis	Vaginal on penis and fingers suspect	ECLI:NL:RBGEL: 2023:2667	Convicted
10	C: Forced oral sex, violent blows on the face victim A: **TRANSFER** from changing balloons during shared nitrous oxide laughing gas use (admits violence)	Saliva on penis suspect; low amounts DNA matching victim	ECLI:NL:RBDHA: 2022:14059	Conviction violence; acquittal sexual abuse
11	C: Sexual assault of 13-yrs old daughter A: Watched porno using a condom, unclear if vaginal cells are from daughter as the mother is not in DNA comparison (**INNOCENT EXPLANATION**)	Vaginal on outer side condom; semen on inner side condom	ECLI:NL:RBOBR: 2016:4724	Convicted
12	C: Sexual assault in asylum seeker center A: Consensual touching (**INNOCENT EXPLANATION**), and suspect did not touch vagina with his finger	Vaginal on fingers suspect	ECLI:NL:RBGEL: 2020:4108	Convicted
13	C: Sexual assaults after knocking victims down on street, RNA typing for case of one victim A: Consensual intercourse (**INNOCENT EXPLANATION**), violence by somebody else	Vaginal in nail dirt suspect	ECLI:NL:RBMNE: 2021:4553	Convicted
14	C: Sexual assault at a train station (digital penetration and touching breasts/ upper body) A: The suspect lifted the victim on a bench (**INNOCENT EXPLANATION**)	Vaginal on fingers (and DNA suspect on upper body victim)	ECLI:NL:RBNHO: 2023:4897	Convicted
15	C: Attempted murder by shooting multiple bullets A: Suspect did not shoot at victim but shot to the ground and car (**INNOCENT EXPLANATION**)	Blood on bullet that went through body with DNA-match victim	ECLI:NL:RBNNE: 2013:5558	Convicted
16	C: Homicide mentally disabled stepdaughter by hitting with baseball bat after repeated violence A: Victim fell from stairs onto baseball bat (**INNOCENT EXPLANATION**)	Blood on baseball bat	ECLI:NL:RBNNE: 2015:656	Convicted

17	C: Murder of business partner by fatal head injuries A: Victim found by suspect who touched him, blood on suspect's clothing from leaning against blooded stair railing (**INNOCENT EXPLANATION**)	Blood, CNS, muscle on clothing suspect; impact pattern bloodstains	ECLI:NL:PHR:2024:90	Convicted
18	C: Fatal stabbing in pub after fighting A: Self-defense, without knife (**INNOCENT EXPLANATION**)	Blood, muscle on knives; no DNA-match suspect on knives	ECLI:NL:GHARL:2021:2893	Acquittal: not certain who stabbed
19	C: Murder partner after inflicting fatal head injuries A: Suspect lifted victim from couch and resuscitated her (**INNOCENT EXPLANATION**)	Blood, muscle, CNS on sweater suspect; impact over contact pattern	ECLI:NL:GHDHA:2021:2295	Convicted
20	C: Fatal shooting with 2 victims A: **SOMEONE ELSE** delivered the fatal shots	Blood, liver, muscle on bullet	ECLI:NL:HR:2022:1591	Convicted
21	C: Suspect shot victim with 6-8 bullets sitting in a car A: It was possible that **SOMEONE ELSE** was wearing the coat (although no statement of lending coat)	CNS on coat suspect (DNA match suspect collar/cuffs coat)	ECLI:NL:GHSHE:2014:4029	Convicted
22	C: Victim beaten on head and died A: **SOMEONE ELSE** wore the clothing	Blood, CNS, muscle on clothing in trash bin	ECLI:NL:RBMNE:2015:7603	Convicted
23	C: Woman in bed attacked by husband, head injuries A: Wounds are self-inflicted, blood pattern is faked by **VICTIM HERSELF** from menstrual/animal blood	Blood on bedding (multiple samples), no menstrual markers	ECLI:NL:GHSHE:2017:2006	Convicted
24	C: Fatal injuries to the head by her partner A: **ANOTHER PERSON** entered the home whilst the suspect was out and killed the victim	Pattern of blood traces on sweater also presence CNS, muscle	ECLI:NL:RBROT:2019:498	Convicted
25	C: Victim is stabbed to death by suspect and expensive watch is taken, found with suspect A: Suspect is not involved, 2 **UNKNOWN MEN** framed him: took his knife and left watch in his car	Muscle and blood on knife blade, mixed DNA profile on knife handle	ECLI:NL:GHARL:2019:2508	Convicted
26	C: Murder by 5 shots in head A: **ANOTHER PERSON** shot; blood on shoes as victim visited suspect and spat to floor with bleeding gums (see ECLI:NL:RBOBR:2019:5920)	Blood, CNS on shoes suspect (and gunshot residues)	ECLI:NL:GHSHE:2023:1845	Convicted
27	C: Forceful insertion ginger as punishment of daughter for inappropriate behavior A: **NO ALTERNATIVE**, daughter is unreliable	Vaginal and blood on pieces of ginger	ECLI:NL:GHAMS:2018:2656	Convicted
28	C: Man found 6-weeks death in bed: suspect confesses homicide: murder or manslaughter A: **NO ALTERNATIVE**, no role other suspect	Blood spatters around bed that fit impact pattern	ECLI:NL:RBOBR:2023:5538	Convicted, **no use BPA or RNA data**
29	C: Son murdered father by 28 stabbings, they co-habited, no traces of another person A: Possibly **ANOTHER PERSON** performed the crime	Blood matching victim on slippers suspect; contact blood trace	ECLI:NL:RBAMS:2016:1839	Convicted, **no use RNA data**
30	C: Murder of husband, no DNA other person than wife on body and clothing victim (17 locations) A: **ANOTHER PERSON** is involved	Vaginal on penis	ECLI:NL:GHARL:2020:9865	Convicted, co-habiting: **not regard** DNA-RNA
31	C: Digital penetration ±10-yrs girl by 82-yrs taxi driver special care transport A: Admits touching body superficially above vagina, no penetration (**LESS INVASIVE EXPLANATION**)	**No indication vaginal** on fingers suspect	ECLI:NL:GHSHE:2024:677	Convicted (touching regarded penetration)
32	C: Exploiting adolescent girls for sex; testing them by digital penetration A: Consensual (**INNOCENT EXPLANATION**)	**No indication vaginal** on fingers suspect	ECLI:NL:RBGEL:2019:2778	Convicted, **no vaginal**[1] no penetration
33	C: Sexual abuse 13-yrs vulnerable girl (multiple times including luring her away from parents) A: Consensual (**INNOCENT EXPLANATION**)	**No RNA results** outer side condom, but DNA matching victim	ECLI:NL:RBDHA:2021:353	Convicted, DNA result suffices

1RNA result is reported as 'indication for' [3]. Since April 2022 based on LR when reporting on vaginal/menstrual [4]

2DNA results match victim except for case 31 in which clothing of the victim was analyzed for DNA matching suspect

Conclusion

RNA typing results can impact court decisions, although when cell type inference does not reflect the central question (e.g. case 28, murder or manslaughter) RNA results are not considered.

Acknowledgments

The authors like to thank all colleagues involved in RNA typing: Natalie de Jong-Weiler, Francisca Duijs, Rachel Rietveld-van Leeuwen, Isabelle Meijssen, and all colleagues involved in reporting of RNA results: Petra Maaskant-van Wijk, Simone van Soest, Judith Gits, Bas Kokshoorn, Martijn Rolloos, Pieterjan Sjoukema.

Conflict of interest statement

No conflict of interest

References

[1] https://uitspraken.rechtspraak.nl/

[2] T. Sijen, S. Harbison, On the Identification of Body Fluids and Tissues: A Crucial Link in the Investigation and Solution of Crime, Genes 12 (2021), Oct 28;12(11):1728

[3] A. Lindenbergh, P. Maaskant, T. Sijen, Implementation of RNA profiling in forensic casework. For Sci Inter:Gen 7 (2013) 159-166

[4] R.J.F. Ypma, P.A. Maaskant-van Wijk, R. Gill, M. Sjerps, M. van den Berge, Calculating LRs for presence of body fluids from mRNA assay data in mixtures. For Sci Int:Gen 52 (2021) 102455

[5] E. Hanson, S. Ingold, G. Dorum, C. Haas, R. Lagace, J. Ballantyne, Assigning forensic body fluids to DNA donors in mixed samples by targeted RNA/DNA deep sequencing of coding region SNPs using ion torrent technology, For Sci Inter:Gen Suppl Series, 7 (2019) 23-24

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 775-781
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Validation of novel mRNA semen-specific markers for forensic fluid identification

D. Borges[1,2], A. Amorim[1,2,3], L. Cainé[4,5], I. Gomes[*1, 3]

[1] *i3S - Instituto de Investigação e Inovação em Saúde Porto, Portugal.*

[2] *FCUP- Department of Biology, Faculty of Sciences of the University of Porto, Portugal.*

[3] *IPATIMUP- Institute of Molecular Pathology and Immunology, University of Porto, Portugal.*

[4] *INMLCF- National Institute of Legal Medicine and Forensic Sciences, I.P., Centre Branch, Portugal*

[5] *FMUP- Faculty of Medicine, University of Porto, EDUM- School of Law, University of Minho, Portugal.*

** Corresponding author at: i3S- Instituto de Investigação e Inovação em Saúde Porto, Portugal; email: igomes@i3s.up.pt*

Abstract

In crime scene investigations, the identification of body fluids using (m)RNA is crucial to provide details about the crime. While DNA profiling gives information on the identity of the donor of a biological stain (by confirming or excluding the donor), RNA analysis, on the contrary, allows for the detection of expressed genes by different cell types. This way the identification of the tissue from which the sample originated, without compromising further DNA analysis (due to co-extraction of RNA and DNA from the same sample), is possible. In the specific context of sexual violence, semen is one of the most relevant fluids. In a scenario where an ejaculation has occurred, the investigation depends mainly on the cytological detection/ visualization of sperm. However, the absence of sperm does not always indicate the absence of semen. The main goal of the present study was to validate four recently and newly described mRNA markers for potential identification of semen in forensic settings. For this, a multiplex system containing the mRNAs (ODF1, SMCP, TcP11 and TNP1) as well as two commonly used markers (PRM1 and SEMG1) was developed. The potential application of this multiplex in samples from forensic backgrounds was evaluated through testing of: i) sensitivity using RNA concentration inputs ranging from 5ng/µl to 0.02 ng/µl; ii) specificity by cross reactivity assessment on vaginal secretions, saliva and menstrual blood samples; and iii) degradation assessment on semen samples up to 18 months old since collection. The preliminary data obtained support the use of the new four mRNA markers in cases of sexual assault. The proposed markers, ODF1, SMCP, TcP11 and TNP1, can serve as a possible molecular alternative for semen identification to the traditional techniques since high sensitivity, high specificity as well as detection in aged samples was confirmed.

Keywords
Body fluid identification, semen, mRNA profiling, forensics

Introduction

In a forensic scenario it is common to find biological stains derived from different or mixed body fluids. Therefore, it is important, not only to identify the donor through DNA profiling, but also to know the source of the sample by identifying the body fluid/ tissue that originated the stain [1-3]. Seminal fluid is one of the most relevant body fluids in the context of sexual violence. In an ejaculation scenario, the investigation depends mainly on the cytological detection of sperm; but the absence of sperm does not necessarily indicate the absence of semen [4]. Also, enzymatic reactions such as the immunochromatographic testing of PSA may lead to a high rate of false positives [5]. The expression of specific messenger RNAs (mRNAs) differ depending on the cell type, so with an mRNA-based assay it is possible to identify the type of biological fluids or cells that make up the tissue using RNA profiling [3]. A biological sample recovered from a crime scene allows, not only for DNA analysis of a genetic profile of an individual, but simultaneously for the identification of the source using RNA analysis [3, 6]. Additionally, reports of high stability of RNA markers with tissue-specific expression patterns have been described, allowing the identification of relevant human cell types in small amounts of samples, as well as in degraded material [7, 8]. This collection of information is crucial to help reconstruct the events of a crime scene and thus for casework conduction since it may offer details about case circumstances [3].

In this study, a hexaplex system was developed for the amplification of six mRNA semen-specific markers in a single PCR reaction which included four novel genes ODF1, SMCP, TcP11 and TNP1 [9], as well as, two commonly used markers, PRM1 [1] and SEMG1 [2].

The application of these markers, particularly, in the investigation of evidence from sexual assaults was evaluated through sensitivity and specificity testing. The stability of the selected markers was also assessed in scenarios of (potentially) degraded samples (approximately up to 18 months of age after collection).

Material and methods

Semen samples: collection, RNA extraction and storage

Fourteen semen samples from male volunteers were collected from healthy, unrelated individuals with ages between 20 and 60 years, under informed signed

consent, and stored at -80ºC. RNA extractions were performed as described in a previous study [10] using the commercial kit ExtractME RNA & DNA (ExtractME, BLIRT S.A., Poland) and manufacturer's recommendations.

Primer design and cDNA synthesis

Gene name, primer sequences, fragment sizes and other relevant information are listed in Table 1. Primer sequences were designed with Primer3 software [11] for the new selected markers (Table 1). The reverse transcription reaction (RT) of the RNA was performed for each sample. cDNA synthesis was performed with the commercial SensiFAST cDNA Synthesis kit (BLIRT S.A., Poland) following the manufacturer's protocol. All cDNA reaction products were diluted in a concentration ratio of 1:10 for subsequent applications.

Table 1. Selected mRNA markers and the respective primer sequences, melting temperatures (Tm), 5´ Fluorocrome and expected amplicon sizes (bps). 6-FAM and YelYak= Freedom™ Dyes (IDT-Integrated DNA Technologies).

Gene	Primer Sequences	Tm*	Lable 5'	Size (bp)	References
SMCP	F: AGGCTGCTGAGGAAGATCAAT	59.17	6-FAM	105	This study
	R: GAACACTTGGAGGTACTGGTC	57.96			
TcP11	F: TCACAGGTCTGACAGAAACCG	59.93	YelYak	117	This study
	R: CCTTGCCTTCCAGACTGCTT	60.25			
ODF1	F: TGCGGCCTGTGTGATCTCTA	60.97	YelYak	143	This study
	R: GTTGTTCTTCTCAGTTTGGCAAG	58.65			
TNP1	F: AAGCCCCTCATTTTGGCAGA	59.89	6-FAM	180	This study
	R: GTAATTGCGATTGGCGTCAT	57.24			
PRM1	F: GCCAGGTACAGTGCTGTCGCAG	65.86	6-FAM	152	[1]
	R: TTAGTGTCTTCTACATCTCGGTCT	58.50			
SEMG1	F: TCGGTAACCATGTGAAAGGA	56.47	6-FAM	120	[2]
	R: GTGTCATCCATGGACCAAGA	57.21			

Multiplex development and optimization: PCR and CE

The multiplex development and optimization resulted in the following optimal PCR conditions in a final volume of 8 µl: 1 µl of primer mix (each primer at 0.25 µM), 4 µl of HotStarTaq Master Mix (Qiagen, Germany) and 2 µl H_2O and 1 µl cDNA (1:10). Thermocycling conditions were: 95°C for 15 minutes; 35 cycles of 94°C for 20 s, 60°C for 30 s, 72°C for 40s; and 72°C for 15 minutes. Capillary

electrophoresis (CE) was performed using GeneScan 500 LIZ dye size internal standard and HiDi formamide in an ABI 3500 Genetic analyzer (AB, Thermo Fisher Scientific, USA). Detection of products was performed using Genemapper v5 (AB, Thermo Fisher Scientific, USA).

Sensibility, specificity and robustness

To address the suitability of the novel semen markers for forensic settings, an RNA sample demonstrating good quality control parameters (high concentration, absence of inhibitors and proteins using the NanoDrop One microvolume UV-Vis spectrophotometer -Thermo Scientific Scientific, USA) was selected. The sample was diluted in a series of 1:4 ratio at a starting point of 5 ng/µl down to 0.02 ng/µl, resulting in five dilution inputs for cDNA synthesis: 5 ng/µl, 1.25 ng/µl, 0.31 ng/µl, 0.08 ng/µl and 0.02 ng/µl.

For the specificity assessment, the six semen-specific markers were also subjected to amplification in different, relevant fluids for forensic casework (particularly sexual assaults) such as (one sample of each body fluid): vaginal secretions, menstrual blood, saliva and a positive control (a semen sample). PCR amplifications of the hexaplex were performed in triplicates.

To infer on robustness and stability of mRNA markers, six semen swab samples were also tested that had been collected at different time points. When RNA extractions were performed, the "age" of the swabs ranged from 502 to 542 days.

Results

a. Multiplex development and optimization

To ensure primer performance, prior to multiplex development, all primer sets were tested in single PCR reactions using standard amplification conditions as in [10]. For the multiplex set-up comprising all six mRNA markers, several optimization procedures were undertaken. The final PCR conditions can be found in the material and methods section and results are shown in Figure 1.

The electropherogram results highlighted the markers expression detecting peaks in a range of approximately 100 bp to 180 bp as expected (example Figure 1). All markers amplified successfully with all six peaks detected in the 14 RNA semen samples tested.

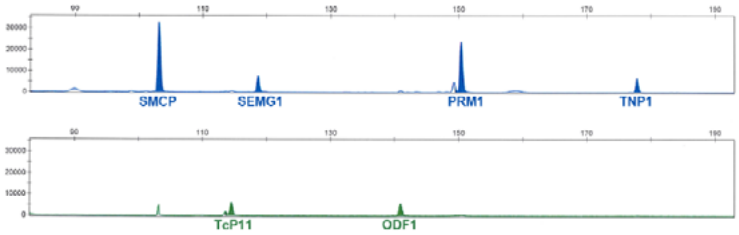


Figure 1. Representative capillary electrophoresis electropherogram of the mRNA semen-specific hexaplex system amplified in one male sample.

b. Sensitivity, specificity and degradation

Regarding the sensitivity results, and as expected, a decrease in peak sizes was observed as the RNA input concentrations decreased. First dropouts started at a dilution point of 0.31 ng/µl for TcP11, for one of the replicas, demonstrating some instability in amplification performance. At 0.08 ng/µl TNP1 and ODF1, in addition to TcP11, also failed to amplify in some of the replicas. At the lowest concentration tested (0.02 ng/µl), no peaks were detected at SEMG1 and TNP1 revealing to be the most sensitive markers. On the other hand, SMCP and PRM1 proved to be more robust due to positive (and still high) detections at 0.02 ng/µl of RNA input in all three PCR replicas tested.

For the specificity testing, no amplification signals were noticed in all of the tested RNA samples from different body fluids, meaning that most likely markers are specific for semen, avoiding cross-reactivity among different tissues. As positive control, a semen sample previously tested was included in the analyses and all six semen-specific markers amplified. These preliminary results show that when analyzing a sample with mixture of body fluids, which is a very common scenario in samples resulting from sexual assaults, the selected six semen-specific markers will only identify semen.

The performance of the hexaplex on potentially degraded samples over time was assessed wherein the collection of samples at time point 1 was 502 days (~17 months) and the older sample was time point 6, 40 days apart, 542 days (~18 months). Time of degradation corresponds to the time since samples were collected till RNA extraction. All markers were still detected even after ~18 months and a decrease of peaks size intensities as well as the presence of a downward slope across the electropherogram (as larger fragments have a lower probability of being amplified) was not observed, confirming the robustness of the studied markers as well as the stability of RNA for forensic settings.

Conclusion

The PCR multiplex was successfully developed based on the expected product sizes for each primer of each mRNA marker that resulted from the "in silico" design. After multiplex optimization, final results showed that the six markers amplified in all samples and a reasonable balance among peaks between all markers was observed. Sensitivity and specificity testing were performed to evaluate the robustness of the mRNA markers for the forensic field. Results have shown that SMCP and PRM1 were the most robust markers, since they were still detected in all three replicas at a sample concentration of 0.02 ng/µl, contrarily to the other markers. On the other hand, specificity results also support that the six markers may be specific for semen since they did not amplify in any of the other different samples tested (vaginal secretions, menstrual blood or saliva). Degradation was also tested in samples aged up to nearly 18 months and all markers were still detected supporting that RNA is more stable than many times considered and therefore can be useful in forensic laboratory settings without major constrains.

It must be highlighted that these represent preliminary data only and more work is needed to validate the specificity of the novel markers (ODF1, SMCP, TcP11 and TNP1) by analyzing more samples from different body fluid types and from different individuals. Also, the multiplex system needs further development by, for example, the inclusion of other markers that may identify other body fluids and tissues specifically relevant for sexual assault cases. The mRNA semen-specific markers evaluated in this study seem to be reliable and stable enough for the identification of semen in the forensic casework.

Acknowledgments

The authors would like to thank Dr. Benedita Ferreira da Silva from INMLCF- National Institute of Legal Medicine and Forensic Sciences, I.P., North Branch, Portugal, for providing the sample swabs of the degradation testing. This work was partially financed by FEDER-COMPETE and by Portuguese funds through FCT- in the framework of the projects "i3S" POCI-01-0145- FEDER-007274. IG is funded by FCT (CEECIND/02609/2017).

Conflict of interest statement

No conflict of interests.

References

[1] Juusola, J. and J. Ballantyne, *Multiplex mRNA profiling for the identification of body fluids.* Forensic Sci Int, 2005. **152**(1): p. 1-12.

[2] Haas, C., et al., *mRNA profiling for the identification of sperm and seminal plasma.* Forensic Science International: Genetics Supplement Series, 2009. **2**(1): p. 534-535.

[3] Sijen, T. and S. Harbison, *On the Identification of Body Fluids and Tissues: A Crucial Link in the Investigation and Solution of Crime.* Genes (Basel), 2021. **12**(11).

[4] Tian, H., et al., *The effect of infertile semen on the mRNA-based body fluid identification.* ELECTROPHORESIS, 2021. **42**(16): p. 1614-1622.

[5] Stroud, A., et al., *A comprehensive study into false positive rates for 'other' biological samples using common presumptive testing methods.* Sci Justice, 2023. **63**(3): p. 414-420.

[6] Lindenbergh, A., P. Maaskant, and T. Sijen, *Implementation of RNA profiling in forensic casework.* Forensic Sci Int Genet, 2013. **7**(1): p. 159-66.

[7] Setzer, M., J. Juusola, and J. Ballantyne, *Recovery and stability of RNA in vaginal swabs and blood, semen, and saliva stains.* J Forensic Sci, 2008. **53**(2): p. 296-305.

[8] Kohlmeier, F. and P.M. Schneider, *Successful mRNA profiling of 23 years old blood stains.* Forensic Sci Int Genet, 2012. **6**(2): p. 274-6.

[9] Salzmann, A.P., et al., *Degradation of human mRNA transcripts over time as an indicator of the time since deposition (TsD) in biological crime scene traces.* Forensic Sci Int Genet, 2021. **53**: p. 102524.

[10] Loureiro, S., et al., *Evaluation of two DNA/RNA co-extraction methods for body fluid identification in forensics.* Forensic Science International: Genetics Supplement Series, 2019. **7**(1): p. 250-252.

[11] Koressaar, T. and M. Remm, *Enhancements and modifications of primer design program Primer3.* Bioinformatics, 2007. **23**(10): p. 1289-91.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 783-788
DOI: https://dx.doi.org/10.15304/cc.2025.1869

The use of rectal mucosa mRNA markers in forensic casework

M. van den Berge[1*], F.E. Duijs[1], P.A. Maaskant – van Wijk[1], T. Sijen[1]

[1]*Netherlands Forensic Institute, The Hague, The Netherlands*

** Corresponding author at: m.van.den.berge@nfi.nl*

Abstract
This study describes the evaluation of rectal mucosa mRNA markers for use in forensic casework. The insight gained from the responses of currently used body fluid mRNA markers in rectal and rectal-related samples highlights the importance of including rectal mucosa-specific markers in forensic mRNA-based body fluid identification assays.

Keywords
Forensic science, mRNA profiling, Body fluid identification, Rectal mucosa.

Introduction

At the Netherlands Forensic Institute, mRNA-based body fluid identification is used to infer the presence of for example vaginal mucosa in sexual assault cases. Recent studies have shown positive responses of genes that have been selected as vaginal mucosa markers, when analysing rectal samples[1-2]. Currently, the mRNA assay used at our institute does not include markers for the identification of rectal mucosa, which may pose a risk for misclassifications when rectal mucosa is present in a trace. To allow for the identification of rectal mucosa, three potential rectal mucosa markers were examined. Additionally, insight was gained in the responses of currently used body fluid mRNA markers by analysing rectal mucosa-containing samples.

2. Materials and methods

Three potential rectal mucosa genes (PHGR1, ZG16 and MUC13) were selected using literature and gene expression database BioGPS [1-3]. Primers were designed using Ensembl and NCBI Primer Blast [4-5]. DNA/RNA co-extraction, DNase treatment, reverse transcription, PCR and CE were performed according to standardized protocols [6].

2.1 Performance on non-target tissues

To investigate specificity of the three above-mentioned rectal mucosa markers, performance of the markers was assessed on different body fluids (blood, saliva, vaginal mucosa, menstrual secretion, semen and nasal mucosa, n=4 each), penile samplings (n=11) and 20 human tissues of the FirstChoice® Human Total RNA Survey Panel (Applied Biosystems) using the input as previously described [6].

2.2 Performance on rectal and rectal-related samples

Rectal mucosa-containing samples were used to assess performance of the rectal mucosa markers on target tissues; and to gain insight in the responses of currently used body fluid mRNA markers [6] in these tissues. These include rectal mucosa (n=8), faeces (n=9) and samples taken around the anus (n=8). Faeces and samples around the anus were included as these samples may contain traces of rectal mucosa.

Rectal mucosa samples were collected using a nylon flocked swab inserted at least five cm into the anus, as the first four cm are anatomically part of the anus where no rectal mucosa is expressed [2]. Faeces samples were collected from the inside of faeces using a nylon flocked swab. Anal samples were taken by swabbing around the anus using a nylon flocked swab. All samples were self-collected with informed consent of voluntary donors.

3. Results and discussion

Primer concentrations were optimized using rectal mucosa RNA extracts of two donors. The optimized primer concentrations were 0.1, 0.1 and 0.15 µM for PHGR1, ZG16 and MUC13, respectively (results not shown).

3.1 Performance on non-target tissues

All three primer sets were included in the Celltyper assay [6] using their optimized primer concentrations. The markers show to be specific, as no rectal mucosa signals were detected in any of the analysed body fluids or penile samplings (**Table 1**). Results of analysing 20 different organ tissues showed detection of PHGR1 and MUC13 in colon and small intestine, which is to be expected as rectal mucosa lines the intestinal wall (results not shown).

Table 1. Marker detection fractions for rectal mucosa markers PHGR1, ZG16 and MUC13 in various body fluids (n=4) and penile skin samples (n=11).

	Blood			Saliva	Saliva /nasal	Nasal	Vaginal mucosa			Menstrual secretion			Semen			Rectal mucosa			Housekeeping	
	HBB	ALAS2	CD93	HTN3	STATH	BPIFA1	MUC4	MYOZ1	CYP2B7P1	MMP10	MMP7	MMP11	PRM1	SEMG1	KLK3	PHGR1	ZG16	MUC13	ACTB	18S-rRNA
Blood	1.00	1.00	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00	1.00
Saliva	0.75	0	0	1.00	1.00	0	0.50	0	0	0	0	0	0	0	0	0	0	0	1.00	1.00
Vaginal mucosa	0.50	0	0.50	0	0	0	1.00	1.00	1.00	0.25	0.25	0	0.50	0.25	0.25	0	0	0	1.00	1.00
Menstrual secretion	1.00	1.00	0.75	0	0	0	1.00	0.50	0.75	1.00	0.75	1.00	0.25	0	0	0	0	0	1.00	1.00
Semen	0.25	0	0	0	0	0.25	0	0	0	0	0	0	1.00	0.75	1.00	0	0	0	1.00	1.00
Nasal mucosa	0.25	0	0.50	0	1.00	0.50	0.75	0	0	0.25	0	0	0	0	0	0	0	0	1.00	1.00
Penile skin	0	0	0	0	0	0	0.18	0	0.09	0	0	0.09	0.27	0.18	0.09	0	0	0	0.64	1.00

3.2 Performance on rectal and rectal-related samples

Figure 1 shows the mRNA profiling results of the Celltyper assay including the three rectal mucosa markers when analysing rectal and rectal-related samples.

Rectal mucosa is "observed" [6] in 75%, 25% and 56% of rectal, anus and faeces samples, respectively (using the x=n/2 guideline: body fluid target marker detection ≥50% is scored "observed"; 0% is "not observed", >0% and <50% is "sporadically observed"). MUC13 appears the least sensitive of the three rectal mucosa markers. This marker is therefore not included in the final assay. The updated version of the Celltyper assay, including rectal mucosa markers PHGR1 and ZG16 is presented in **Table 2**.

Non-target signals are detected throughout the three different sample types, mainly for the different vaginal mucosa markers. Using the x=n/2 guidelines, "observed" scorings would be obtained for vaginal mucosa in two of the rectal mucosa samples and one of the faeces samples. The two rectal samples resulting in an "observed" scoring for vaginal mucosa were both samples from male donors (**Figure 1A**).

For each of the samples, the LR for the presence of vaginal mucosa and/or menstrual cells was calculated using the RNA LR tool [7]. For the two male rectal mucosa samples with an "observed" scoring for vaginal mucosa, LRs >200 were

obtained (**Figure 1A**). These LRs fall in the highest verbal scale that is currently used for the reporting of RNA LRs at our institute [7] (LR values 100-1000; "appreciably more probable"), indicating that if no testing is done for the presence of rectal mucosa, LRs can be erroneously supportive for the presence of vaginal and/or menstrual secretion in rectal mucosa-containing samples.

At this moment, the RNA LR tool lacks knowledge on rectal mucosa markers, and its dataset does not include rectal mucosa(-containing) samples. As a result, the tool is unable to recognise rectal mucosa samples, or account for the effects of marker responses of currently used body fluid markers in rectal mucosa-containing samples. To address this, the dataset needs to be updated to include this information. Therefore, at this time, no statement can be made regarding the presence of vaginal and/or menstrual secretion when rectal mucosa is scored "observed" using the x=n/2 guidelines.

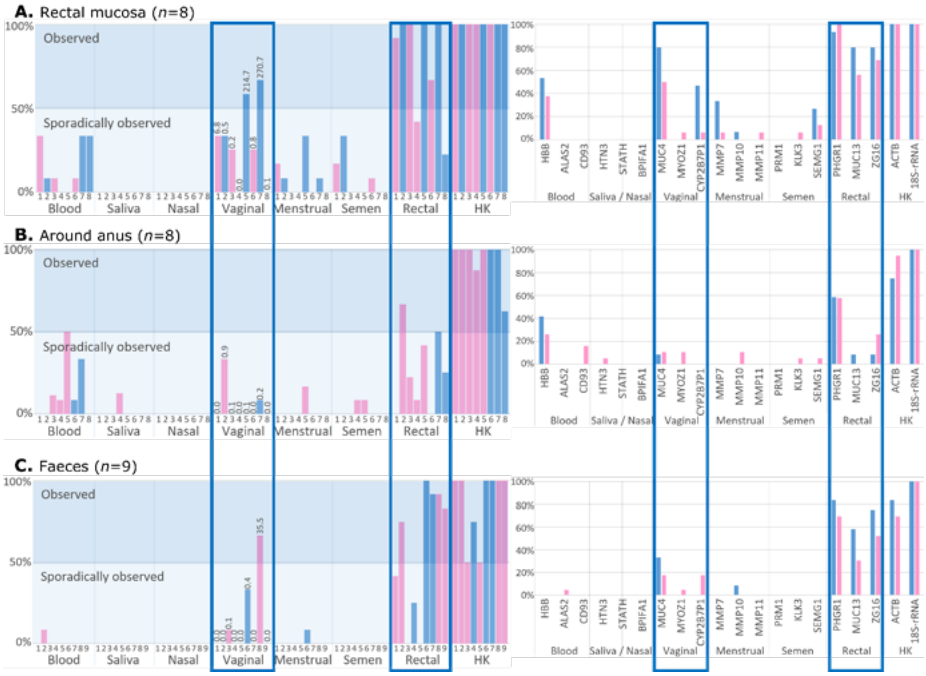


Figure 1. Left: Per body fluid per donor the scoring results using x=n/2 guidelines [6]. The LR for vaginal and /or menstrual cells in these samples is presented above the vaginal bars [7]. **Right**: detection % per marker. Data is shown separately for male (blue bars) and female (pink bars) donors. HK=housekeeping.

Table 2. Primer sequences in the updated Celltyper multiplex, including rectal mucosa markers PHGR1 and ZG16.

Gene	Tissue	[primer] µM	Forward primer (5'-3') Reverse primer (5'-3')	Size (bp)	Dye+	Ref
HBB	Blood	0.035	GCACGTGGATCCTGAGAACTTCAG ATGGGCCAGCACACAGACCAG	61	6-FAM™	[6]
CD93	Blood	0.075	GCTCTGGGGCTACTGGTCTATC TCCCAGGTGTCGGACTGTACTG	151	NED™	[6]
ALAS2	Blood	0.05	TTCTGCACCAGAAGGACTCAGCC* TAAATCTCGCACCCTGGCAGGATC	103	6-FAM™	[6]
HTN3	Saliva	0.2	CTTCACTTCAGCTTCACTGACTTCTG CTTTGCATGTGAATCAGCTCCAGTC	132	VIC®	[6]
STATH	Saliva/nasal mucosa	0.3	TTCATCTTGGCTCTCATGGTTTCCATG GCCATACCCATAACCGAATCTTCCA	93	6-FAM™	[6]
BPIFA1	Nasal mucosa	0.15	CAAGTGAATACGCCCCTGGTCG GAATGGGTGCAGTCACCAAGGAC	131	PET™	[6]
SEMG1	Seminal fluid	0.6	GGAAGATGACAGTGATCGT CAACTGACACCTTGATATTGG	121	6-FAM™	[6]
KLK3	Seminal fluid	0.05	GACGTGGATTGGTGCTGCACC CTTCTCGCACTCCCAGCCTC	64	PET™	[6]
PRM1	Spermatozoa	0.05	GAGAGCCATGAGGTGCTGCC AGGCAGGAGTTTGGTGGATGTGC	90	NED™	[6]
MUC4	Vaginal mucosa	0.8	CTGCTACAATCAAGGCCACTGCTAC AAGGGAAGTTCTAGGTTGACAGTTGG	141	6-FAM™	[6]
CYP2B7P1	Vaginal mucosa	0.05	CCTCATGTCGCAGAGAGAGTCTAC CCCATGGGGAGAAGGTCAGCA	146	VIC®	[6]
MYOZ1	Vaginal mucosa	0.8	CGTGTTCTCCGGTCACAGCAG TGGATTCAGCCGGCTGCTCG	88	VIC®	[6]
MMP7	Menstrual secretion	0.4	GAACAGGCTCAGGACTATCTC TTAACATTCCAGTTATAGGTAGGCC*	127	VIC®	[6]
MMP10	Menstrual secretion	0.07	GCATCTTGCATTCCTTGTGCTGTTG GGTATTGCTGGGCAAGATCCTTGTT	107	VIC®	[6]
MMP11	Menstrual secretion	0.15	CAACCGACAGAAGAGGTTCG GAACCGAAGGATCCTGTAGG	76	NED™	[6]
PHGR1	Rectal mucosa#	0.1	CCAAAAAGATGGACCCAGGTCCG AGCAGTGACCTGGAGGATGGC	72	6-FAM™	This paper
ZG16	Rectal mucosa#	0.1	GGGCTGCTTTGCATCTGAAACTGT TCCATACTCTCCACTATAGGAGGAAGAC	127	6-FAM™	This paper
ACTB	Housekeeping	0.0225	CAGAGCCTCGCCTTTGCCGAT CGCGGCGATATCATCATCCATGGT	75	PET™	[6]
18S-rRNA	Housekeeping	0.013	GACTCAACACGGGAAACCTCACC CTCCACCAACTAAGAACGGCCATG	110	PET™	[6]

+Forward primers are labelled. *Underlined nucleotides are 5' tails added to improve multiplex spacing.
#Rectal marker MUC13 was not included in the final assay (forward: GACCCAGAAGAGAAACATTCCATGGC, reverse: TGGTGACAGAGATGTGCTTACAGTAAGAA, NEDTM, 132 bp)

Concluding remarks

Results highlight the importance of introducing rectal mucosa-specific markers to currently used assays, to avoid the risk of falsely concluding the presence of vaginal mucosa in samples containing rectal mucosa.

The addition of these markers allows for the evaluation of faecal matter or rectal mucosa, allowing to answer casework questions that could previously not be answered by use of mRNA-profiling.

Acknowledgments

We are grateful for all the donors who contributed samples for this study.

Conflict of interest statement

None

References

[1] T. Akutsu, et al., Evaluation and simultaneous determination of rectal mucosa markers by multiplex reverse transcription-PCR for biological evidence of sexual assault with anal penetration. Forensic Sci. Int. Genet. 59 (2022) 102712.

[2] M. Bamberg, et al., Development and validation of an mRNA-based multiplex body fluid identification workflow and a rectal mucosa marker pilot study, Forensic Sci. Int. Genet. 54 (2021) 102542.

[3] C. Wu, et al., BioGPS: building your own mash-up of gene annotations and expression profiles. Nucleic acids research, (2016) 44(D1), D313-D316.

[4] P. Flicek, et al., Ensembl 2014, Nucleic Acids Res. 42 (2014) D749–D755.

[5] J. Ye, et al., Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, BMC Bioinf. 13 (2012) 134.

[6] M. van den Berge, et al., Extended specificity studies of mRNA assays used to infer human organ tissues and body fluids. Electrophoresis 2017, 38, 3155–3160.

[7] R.J.F. Ypma, et al., Calculating LRs for presence of body fluids from mRNA assay data in mixtures. For Sci Int:Gen 52 (2021) 102455.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 789-796
DOI: https://dx.doi.org/10.15304/cc.2025.1869

New miRNA markers to distinguish saliva from vaginal secretion in body fluid source identification

Xunhong Yang[1], Zhilong Li[2], Jing Zhu[3], Li Wang[4*]

[1]*West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu, Sichuan, 610041, P.R. China.*

[2]*Laboratory of Molecular Translational Medicine, Center for Translational Medicine, Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University), Ministry of Education, West China Second University Hospital, Sichuan University, Chengdu, Sichuan,610041, P.R. China.*

[3]*Forensic Science and Technology Department, Sichuan Police College, Luzhou, Sichuan, 646000, P.R. China.*

[4]*Department of Medical Genetics, Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University), West China Second University Hospital, Sichuan University, Chengdu, Sichuan, 610041, P.R. China.*

Corresponding author at: No. 20, Section 3, South Renmin Road, Chengdu, Sichuan Province, 610041, P.R. China, wicky.wangli@qq.com, +86 (028) 85501269.

Abstract

Several studies have reported that miRNA can be used for forensic body fluid identification, and a series of body fluid-specific miRNA have been proposed. However, distinguishing between saliva and vaginal secretions remains a challenge. This study focused on identifying novel miRNA markers to improve the differentiation between saliva and vaginal secretions in forensic body fluid identification. A panel of three previously unreported miRNAs was identified using Arraystar small RNA microarray, showing significant differences in expression between the two body fluids. U6-snRNA was used as the reference gene, with both an old and a new version tested. The results indicated that using different versions of U6-snRNA led to contrasting outcomes. The three new miRNAs, together with our previously identified 12 loci, form a 15-miRNA body fluid-specific subset. Using the MLR method to model with data from 15 loci, the LOOCV accuracy reached 100% when using the old version U6-snRNA as the reference, and 98% when using the new version U6-snRNA as the reference. With new U6-snRNA as a reference, the simplified four-loci model achieved a test set accuracy of 83.33%. Our study identified three novel miRNA loci that can be used to distinguish epithelial cell origins, while also highlighting the importance of reference selection in qPCR.

Keywords
Forensic genetics, body fluid identification, miRNA, multiple logistic regression, qPCR analysis.

Introduction

Identifying the type and origin of biological samples left at a crime scene is crucial in forensic investigations, as it can provide clues for crime scene reconstruction and linkages between victim, perpetrator, and scene. Existing presumptive tests for body fluid identification in forensics are mainly based on chemical, enzymatic, or immunological reactions [1], which are prone to various limitations, inability to positively confirm the presence of certain biological fluids, consumption of valuable samples, and varying degrees of sensitivity and specificity. Advances in human genomics and molecular genetics have provided us with new choices, such as DNA methylation, mRNA, and microRNA (miRNA) [2]. Compared with traditional methods, these markers permit parallel analysis of different biological fluids.

For RNA, there is no unnecessary sample consumption as it can be simultaneously extracted from the specimen [3, 4]. Compared to mRNAs, miRNAs are considered more stable due to their small size and protection from nucleases [5, 6]. In addition, miRNAs present cell-specific expression patterns [7], which are considered as ideal markers to distinguish different body fluids.

Earlier, we developed a panel of 12 target miRNAs to differentiate between five types of body fluids [8]. By integrating this with a machine learning algorithm, the predictive models achieved 100% accuracy. However, despite the strong performance, the model lacked miRNAs specific to vaginal secretions, making it challenging to simplify the model. In the current study, three saliva samples and three vaginal secretion samples were analyzed using the Arraystar small RNA microarray, which simultaneously detects tsRNAs and miRNAs within a single array. Three new miRNAs were screened out to be vaginal-specific, and the expression of three miRNAs showed significant differences between saliva and vaginal secretions using the miRCURY LNA miRNA PCR System. Two versions of U6-snRNA were used as the reference respectively. A machine learning method, multiple logistic regression (MLR), was used for modeling, and attempts were made to simplify and optimize the models to better suit forensic practice.

Materials and methods

Sample collection

The samples used for chip detection were newly collected, while the samples for subsequent reverse transcription quantitative polymerase chain (RT-qPCR) were from the same batch as those used in previous study [8]. A total of 83 samples, containing venous blood, semen, semen-free vaginal secretions, menstrual blood and saliva, were collected from healthy, unrelated individuals with the approval of the Ethics Committee at West China Second University Hospital, Sichuan University. Written informed consent was obtained from all participants, which included 26 males and 30 females, aged 21 to 45 years. This study adhered to the ethical guidelines outlined in the World Medical Association's Helsinki Declaration [9]. All samples were stored at –80 °C until RNA extraction. Three newly collected saliva samples and three vaginal secretion samples were used for chip detection, while the 62 single-source body fluid samples from the original study were used for quantitative validation.

RNA Isolation and Quality Control

For array detection, total RNA was isolated from approximately 2 mL of saliva and a full vaginal secretion swab using TRIzol (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The RNA quantity was measured with a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). For subsequent RT-qPCR, total RNA of liquid samples of saliva (2 mL), peripheral blood (0.5 mL), semen (0.5 mL), half of a vaginal secretion swab, and menstrual blood gauze (approximately 2 cm^2) were extracted using the RNeasy Mini Kit (Qiagen, Hilden, NRW, Germany) according to the manufacturer's instructions. The total RNA integrity extracted from body fluid stain samples was generally low; however, both literature reports [10, 11] and our own experience suggest that this does not affect chip detection and miRNA quantification. Therefore, RNA integrity of the samples was not assessed in this study.

Array Analysis and Candidate Selection

The Small RNA Microarray (Arraystar, Rockville, Maryland, USA) utilized in this study contains over 14,000 probes designed to specifically quantify various small RNA biotypes. In summary, 100 ng of total RNA was used from each sample, and dephosphorylated to create a uniform 3'-OH end. The RNA with 3'-OH ends was denatured using DMSO and subsequently labeled enzymatically with Cy3. The

Cy3-labeled small RNAs were hybridized to Arraystar Human Small RNA Arrays using an Agilent Hybridization Oven (Agilent Technologies, Santa Clara, CA, USA). The microarray scanning and data analysis were performed by KangChen Bio-tech (Shanghai, China). Selection of candidate sites was based on the array results, focusing on miRNA, and ranked according to fold changes (FC) from highest to lowest. Statistical significance (p-value) and signal intensity were also considered. Sites showing high FC but low signal intensity were deemed difficult to detect.

RT-qPCR

To assess the relative expression of miRNA candidates, the miRCURY LNA miRNA PCR System (Qiagen, Hilden, NRW, Germany) was utilized. Total RNA was diluted to a concentration of 10 ng/μL with nuclease-free water. Complementary DNA (cDNA) was synthesized from 20 ng of total RNA using the miRCURY LNA RT Kit (Qiagen, Hilden, NRW, Germany) following the manufacturer's instructions. Reverse transcription was performed using a PCR thermal cycle instrument (Eppendorf, Hamburg, Germany) with the following conditions: 42°C for 60 minutes, 95°C for 5 minutes, followed by holding at 4°C. An RT-negative control (without reverse transcriptase) was included to rule out potential genomic DNA contamination. The 12 samples used in the subsequent evaluation test set were renamed a–l for this procedure. The resulting cDNA was then diluted 1:8 for further analysis. An amplification mixture was prepared according to the miRCURY LNA SYBR® Green PCR Kit (Qiagen, Hilden, NRW, Germany) protocol. All miRCURY miRNA PCR assays were performed in triplicate using a cobas 4800 Real-Time Quantitative Analysis System (Roche Diagnostics, Basel, Switzerland). The PCR conditions were set to 95°C for 2 minutes, followed by 40 cycles of 95°C for 10 seconds and 56°C for 1 minute. In the negative control, cDNA was replaced with an equal volume of water to detect any potential contamination. All experiments were conducted in full accordance with the essential information outlined in the MIQE guidelines [12], ensuring the reliable and unambiguous interpretation of qPCR results.

Statistical analysis

Relative expression of miRNA was calculated using the delta (ΔCq) method: $\Delta Cq = Cq_{\text{Target miRNA}} - Cq_{\text{U6-snRNA}}$. Normal distribution was checked using SPSS 26.0 (IBM, USA). The expression difference between saliva and vaginal secretion was analyzed by independent samples t-test (for normally distributed data) or non-parametric tests (for skewed distributions). Differences in miRNA expression across the five types of body fluids were assessed using ANOVA (for normally distributed data) or non-parametric tests (for skewed distributions).

Construction and simplification of the predictive model

First, modeling was performed using all 15 parameters, and the Leave-One-Out Cross-Validation (LOOCV) and test set accuracy were obtained. Then the MLR model was employed to reduce the number of parameters for modeling. L1 regularization was applied to perform feature selection, progressively reducing the number of parameters from 15 to 4. After each reduction step, LOOCV and test set accuracy were evaluated to assess model performance. The above operations were all performed on the Python platform.

Results

Array experiments and candidates' selection

All samples used in the array screening experiment successfully passed quality control, and RNA extraction for the subsequent RT-qPCR was successful. The A260/A280 ratios ranged from 1.80 to 2.1. Samples detected using RT-qPCR here were the same batch as those used in the previous article [8]. Hsa-miR-26a-5p, hsa-miR-762 and hsa-let-7a-5p were screened out based on the array results. The chip results showed that miR-762 was highly abundant in vaginal secretions, with a fold change of 71.13 compared to saliva samples, while let-7a-5p and miR-26a-3p were highly expressed in saliva, with fold changes of 34.62 and 41.85 compared to vaginal secretions, respectively. Two versions U6-snRNA were used as the references.

Expression of new miRNAs across five body fluids

Our previous research suggested that different reference selections might influence the specificity of miRNAs. In this study, we used both the new version (V2) of U6-snRNA and the old version (V1) of U6-snRNA as the single reference. The V2 U6-snRNA showed higher expression than the V1 U6-snRNA in the same samples. When the V1 U6-snRNA was used as the reference, the expression of three miRNAs was higher in vaginal secretion samples compared to saliva, whereas the opposite result was observed when the V2 U6-snRNA was used. The two versions used here are from the same system but have different sequences, further highlighting the importance of reference selection in qPCR.

Performance of MLR model for miRNA data analysis

In the test set, which included 12 single-source body fluid samples, we assessed both LOOCV accuracy and test set accuracy as the number of parameters

was progressively reduced, using V1 U6-snRNA and V2 U6-snRNA as reference, respectively. Based on the results of both LOOCV and test set accuracy, we selected the seven-parameter model and the four-parameter model as simplified predictive models. When using V1 U6-snRNA as the reference and 15 loci, modeling with 50 samples achieved a LOOCV accuracy of 96%. After reducing the number of loci to seven, using logistic regression with an L1 penalty, the LOOCV accuracy improved to 98% (Figure 1A). The simplified seven-loci model included miR-144-3p, miR-141-3p, miR-891a-5p, miR-223-3p, miR-205-5p, miR-1246, and miR-1260b, achieving a test set accuracy of 100%.

Similarly, with V2 U6-snRNA as the reference, the full-parameter model reached a LOOCV accuracy of 98%, and reducing the parameters to four resulted in a LOOCV accuracy of 96% (Figure 1B). The simplified four-loci model, with miR-144-3p, miR-223-3p, miR-205-5p, and miR-762, achieved a test set accuracy of 83.33%. In this model, one saliva sample was misclassified as vaginal secretion, and one vaginal secretion sample was misclassified as saliva.

Discussion and conclusion

In this study, we used microarray screening to identify three previously unreported miRNAs for distinguishing between saliva and vaginal secretion samples. The results showed that using different versions of U6-snRNA as a reference led to different, and sometimes even opposite, outcomes. By combining the previous 12 miRNAs and modeling with 15 miRNAs using MLR, the model achieved 100% accuracy in the test set. When the number of loci was reduced to eight, the LOOCV accuracy reached 96% (new version U6-snRNA as a reference) and 100% (old version U6-snRNA as a reference), respectively. The results of this study further confirm the potential of miRNA in forensic body fluid identification, demonstrating that, when combined with machine learning, a limited number of miRNAs can achieve high predictive accuracy.

Acknowledgments

None.

Conflict of interest statement

The authors declare no conflict of interest.

Funding

This work was supported by Sichuan Science and Technology Program (No. MZGC20230066).

References

[1] C.L. GLYNN, Potential applications of microRNA profiling to forensic investigations, RNA (2020) 1-9.

[2] M. Kayser, P. de Knijff, Improving human forensics through advances in genetics, genomics and molecular biology, Nature Reviews Genetics 12(3) (2011) 179-192.

[3] Y. Li, J. Zhang, W. Wei, et al. A strategy for co-analysis of microRNAs and DNA, Forensic Science International: Genetics 12 (2014) 24-29.

[4] D. van der Meer, M.L. Uchimoto, G. Williams, Simultaneous Analysis of MicroRNA and DNA for Determining the Body Fluid Origin of DNA Profiles, Journal of Forensic Sciences 58(4) (2013) 967-971.

[5] J.D. Arroyo, J.R. Chevillet, E.M. Kroh, et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma, Proceedings of the National Academy of Sciences 108(12) (2011) 5003-5008.

[6] K.C. Vickers, B.T. Palmisano, B.M. Shoucri, et al. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins, Nat Cell Biol 13(4) (2011) 423-33.

[7] P. Landgraf, M. Rusu, R. Sheridan, et al. A mammalian microRNA expression atlas based on small RNA library sequencing, Cell 129(7) (2007) 1401-14.

[8] Z. Li, M. Lv, D. Peng, et al. Feasibility of using probabilistic methods to analyse microRNA quantitative data in forensically relevant body fluids: a proof-of-principle study, Int J Legal Med 135(6) (2021) 2247-2261.

[9] W.M. Association, World Medical Association Declaration of Helsinki Ethical Principles for Medical Research Involving Human Subjects, JAMA 310 (2013) 2191-2194.

[10] M. Jung, A. Schaefer, I. Steiner, et al. Robust microRNA stability in degraded RNA preparations from human tissue and cell samples, Clin Chem 56(6) (2010) 998-1006.

[11] J.S. Hall, J. Taylor, H.R. Valentine, et al. Enhanced stability of microRNA expression facilitates classification of FFPE tumour samples exhibiting near total mRNA degradation, Br J Cancer 107(4) (2012) 684-94.

[12] C.T. Wittwer, J. Vandesompele, G.L. Shipley et al. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments, Clinical Chemistry 55(4) (2009) 611-622.

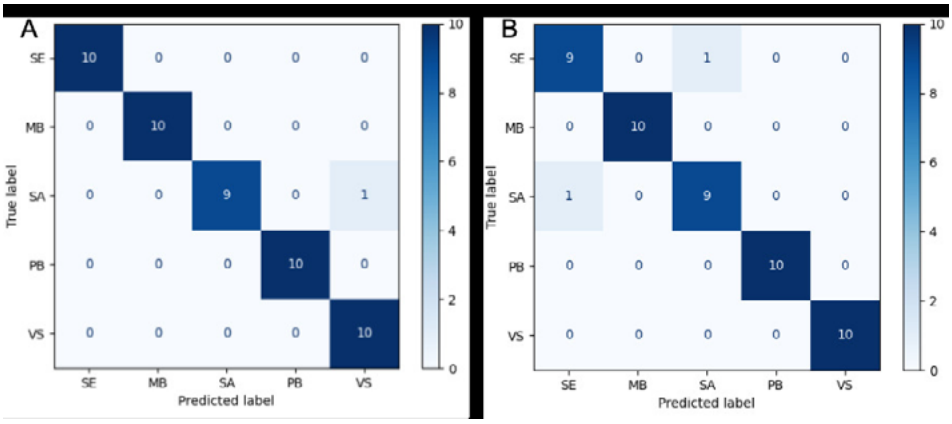


Figure 1. LOOCV confusion matrix for the simplified model. A, V1 U6-snRNA used as a reference (seven-loci model). B, V2 U6-snRNA used as a reference (four-loci model).

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 797-801
DOI: https://dx.doi.org/10.15304/cc.2025.1869

DNA methylation direct sequencing in body fluid identification

Tangleixin Li[1+], Yuanyuan Xiao[1+], Haoyan Jiang[1], Yuxuan Tan[1], Meili Lv[2], Yuan Li[1], Shengqiu Qu[1], Lin Zhang[1], Weibo Liang[1*], Bing Long[3, 4*]

[1]*Department of Forensic Genetics, West China School of Basic Medical Sciences and Forensic Medicine, Sichuan University, Chengdu, 610041, Sichuan, P.R. China*

[2]*Department of Immunology, West China School of Basic Medical Sciences and Forensic Medicine, Sichuan University, Chengdu, 610041, Sichuan, P.R. China*

[3]*Sichuan Police College, Luzhou, Sichuan, 646000, P.R. China*

[4]*National Engineering Laboratory for Forensic Science, Institute of Forensic Science, Ministry of Public Security, No.17 Muxidi Nanli, Beijing 100038, China*

[+]*These authors contributed equally to this work.*

** Correspondence authors: Bing Long 24691982@qq.com, Weibo Liang liangweibo@scu.edu.cn.*

Abstract

The detection of DNA methylation with the typical method has to be applied in forensic genetics for body fluid identification, disease prediction, age estimation and so on [1,2]. However, the typical bisulfite conversion method always cannot avoid high sample loss and sample degradation due to the stringent reaction conditions. To avoid these problems and detect DNA methylation correctly and efficiently, the newly single-molecule sequencing has been used in this study [3]. Here, DNA methylation was performed on genomic DNA (gDNA) and amplified DNA based on QNome nanopore sequencing technology. The preliminary target region for detection consists of 16 body fluid-specific CpGs in our previous research. Then we compared the results obtained from QNome sequencing and Capillary Electrophoresis (CE) platform to evaluate the accuracy of single-molecule sequencing. The results showed that the methylation levels of the two types of samples exhibited consistent trends. Therefore, the results support that combined single-molecule sequencing with DNA methylation would prompt the identification of body fluid and even enable differentiate a greater variety of body fluid by detecting larger DNA methylation.

Keyword

DNA methylation, single-molecule sequencing, body fluid identification

Introduction

DNA methylation has attracted much attention from researchers because of its involvement of gene expressions and life development. Since DNA methylation was discovered from 1950s, it has been applied in many fields, such as the identification of body fluid, age estimation, facial feature prediction, clinical disease prediction and so on. Typically, to be detected, the DNA methylation has to be converted by the bisulfite treatment. Although this method was considered as the gold standard, it has disadvantages of sample losses and degradation. The emergence of single-molecule sequencing enables direct detection of the methylation of gDNA without additional sample processing, avoiding sample loss, and offering higher throughput [4]. It's a real-time sequencing technology and does not require a pause between read steps, has the advantage of providing longer read lengths, kinetic variation information, and shorter run times. Therefore, this study aims to use single-molecule sequencing technology to detect DNA methylation and preliminarily explore its application in forensic body fluid identification. We initially screened 16 body fluid-specific CpGs, which were selected from our previous research. Capillary electrophoresis (CE) and single-molecule sequencing technology were employed to simultaneously detect DNA templates before and after amplification, and we compared the results to explore the accuracy and applicability of single-molecule sequencing technology.

Materials and methods

Semen DNA is extracted using the QIAmp DNA Mini kit (Qiagen, Hilden, Germany) and quantified with both Nanodrop and the Qubit dsDNA HS assay kit (Thermo Fisher Scientific, USA). Subsequently, semen DNA was amplified using our previously developed 16 CpGs-panel and then subjected to single-molecule sequencing and CE, while it will also undergo single-molecule sequencing. The library is then prepared using the QDL-E v2.0 kit, and sequencing is performed on the QNome-3841 sequencer. For methylation analysis, single-molecule sequencing data is processed using software Minimap2, Samtools, and f5c. Finally, we compared the methylation results based on different platforms.

Results and discussion

The methylation results of gDNA and amplified DNA were compared with the results obtained from CE platform. Inferred base accuracy is approximately around 80% and 90% respectively, the raw reads were filtered by NanoFilt with Q-value > 7 and Q-value >12 (Table 1). The count of unmethy- and methy- reads of target CpGs and the calculated methylation ratio by measuring methylation signals on reads were shown in Figure 1 and 2. Compared the methylation levels of PCR products and gDNA detected on different platforms, we can indicate that the two types of samples exhibited the consistent trends of methylation levels, but the low sequencing depth due to low throughput resulted in significant differences in CpG methylation levels. PCR samples were analyzed using Q7 and Q12 data, with similar results under both filtering conditions. The analysis results with Q12 data were closer to the reference CE results (Figure 3). Through the results, we can see the potential of the combination of DNA methylation and single-molecule sequencing technology for body fluid identification.

Table 1. The relationship between sequencing quality scores and inferred base accuracy

Quality Score	Probability of detecting erroneous bases	Accuracy of inferred base detection
Q7	≈1/5	≈80%
Q10	1/10	90%
Q20	1/100	99%
Q30	1/1000	99.90%

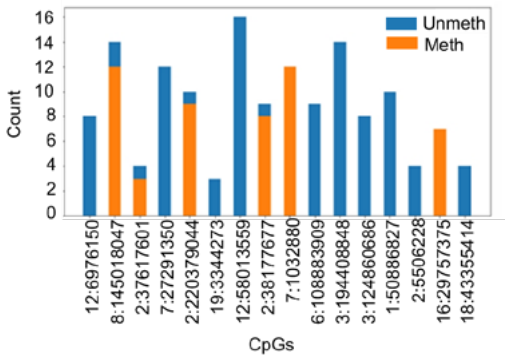


Figure 1 The count of unmethy- and methy- reads of target CpGs. The X-axis represents the physical location of CpGs. The Y-axis represents the count of unmethy- and methy- reads of target CpGs.

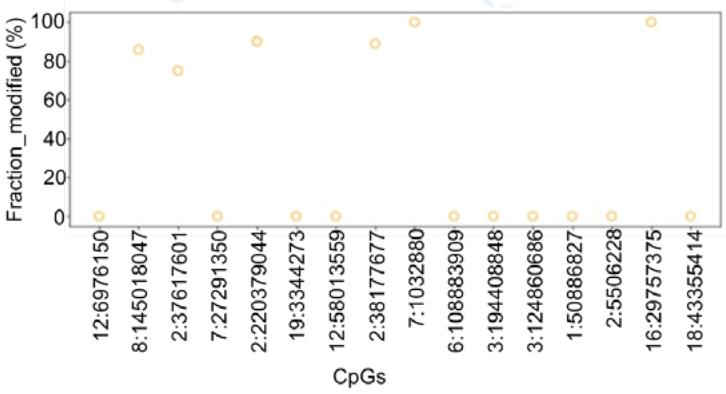


Figure 2 The methylation ratio of target CpGs. The X-axis represents the physical location of CpGs. The Y-axis represents the percentage of fraction-modified of target CpGs.

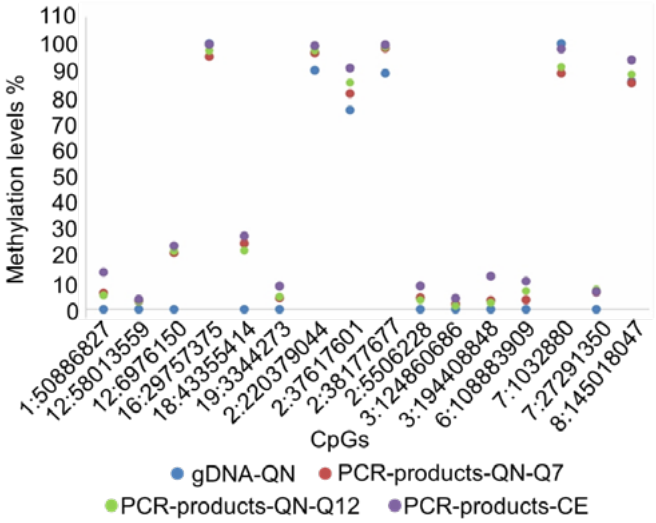


Figure 3 Comparison of methylation levels between PCR products and gDNA detected on different platforms. The X-axis represents the physical location of CpGs. The Y-axis represents the percentage of the methylation levels of target CpGs. The blue points represent the results of gDNA that detected by single-molecule sequencing. The red points represent the results of PCR products that filtered by Q7 after detection by single-molecule sequencing. The green points represent the results of PCR products that filtered by Q12 after detection by single-molecule sequencing. The purple points represent the results of PCR products that detected by CE.

Conclusion

This preliminary study demonstrates that single-molecule sequencing technology can accurately detect methylation levels, showing potential for application in forensic body fluid identification. In the future, it is hoped that detection of larger DNA methylation regions will enable differentiation of a greater variety of body fluids.

Acknowledgement

We would like to thank all the participants who volunteered for this study. TLXL drafted and proofread the manuscript. YYX performed the experiments and data analysis, and reviewed the manuscript. HYJ and YXT contributed to the sample collection and experiment. YL and MLL contributed to the experiment and supervised the project. BL contributed to the sample collection and the optimization of the experiment. SQQ and WBL conceived and supervised the project, and designed the study. LZ supervised the project. All the authors have read and approved the final version of the manuscript. This work was supported by the Open Project of National Engineering Laboratory for Forensic Science (2021NELKFKTO5).

Conflict of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Lee, H.Y., et al., Potential forensic application of DNA methylation profiling to body fluid identification. Int J Legal Med, 2012. 126(1): p. 55-62.

[2] Noroozi, R., et al., DNA methylation-based age clocks: From age prediction to age reversion. Ageing Res Rev, 2021. 68: p. 101314.

[3] Rhoads, A. and K.F. Au, PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics, 2015. 13(5): p. 278-89.

[4] Branton, D., et al., The potential and challenges of nanopore sequencing. Nat Biotechnol, 2008. 26(10): p. 1146-53.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 803-808
DOI: https://dx.doi.org/10.15304/cc.2025.1869

The potential application of Enzyme conversion of DNA methylation in body fluid identification

Yihang Huang[a,1], Yuanyuan Xiao[a,1], Manrui Li[a], Haoyan Jiang[a], Yuxuan Tan[a], Xiameng Chen[a], Shengqiu Qu[a], Weibo Liang[a,*], Peng Bai[a,*]
Yihang Huang and Yuanyuan Xiao contributed equally to this work.

[a] *Department of Forensic Genetics, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu, 610041, Sichuan, P.R. China*

**Correspondence author: Weibo Liang liangweibo@scu.edu.cn, Peng Bai baipeng@scu.edu.cn.*

Abstract
DNA methylation regulates gene expression and plays significant roles in mammals' development and disease processes, while also having widely applications in forensic science. This study investigates a new enzymatic conversion method for DNA methylation analysis, which utilizes TET2 and APOBEC enzymes to convert the unmethylated cytosine to uracil while methylated cytosine unchanged. In addition, the method is compared with the widely used bisulfite conversion method in the context of forensic body fluid identification (BFI). A total of 100 samples from diverse body fluids were examined for methylation levels at 16 body fluid-specific CpGs. The average conversion efficiency and standard deviation for enzymatic conversion were 0.841±0.079, and stable results were obtained at low DNA input. Notably, at DNA inputs of ≤10 ng, bisulfite conversion tended to overestimate methylation values, and random CpG drop-out was observed in both methods. These results support the potential of enzymatic conversion as a promising alternative for improving accuracy and reliability in BFI within forensic applications.

Keywords
DNA methylation; body fluid identification; enzymatic conversion; forensic genetics

1. Introduction

DNA methylation is crucial for regulating gene expression and has extensive applications in forensic science, including BFI, age estimation, and differentiation of identical twins. The common conversion method for DNA methylation is bisulfite

treatment, which involves harsh reaction conditions and can result in sample loss and DNA degradation. Recently, researchers have developed an enzymatic conversion method that uses TET2 to oxidize 5-methylcytosine to 5-carboxycytosine, followed by APOBEC-mediated deamination of unmethylated cytosines into uracil [1]. The sequence generated by enzymatic conversion is identical to that produced by bisulfite conversion. This method operates under mild reaction conditions, offers high conversion efficiency with minimal sample loss [2,3], making it a promising method for challenging forensic samples. This study aims to evaluate the application of enzymatic conversion for forensic BFI and compare it with the bisulfite conversion.

2. Materials and methods

In this study, a total of 100 samples were collected, comprising 20 samples each of peripheral blood (BL), semen (SE), menstrual blood (MB), saliva (SA), and vaginal secretions (VS). We selected 16 body fluid-specific CpGs to evaluate the conversion methods in this study [4]. DNA extraction was carried out with the QIAmp DNA Mini Kit (Qiagen, Hilden, Germany). For bisulfite conversion, the Epitect® Bisulfite Kit (QIAGEN, Hilden, Germany) was utilized, while enzymatic conversion was performed using the NEBNext® Enzymatic Methyl-seq Conversion Module (New England Biolabs, USA). Additionally, DNA quantification was conducted using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, USA). We performed quantitative PCR (qPCR) analysis on the converted DNA samples to assess the conversion efficiency of the two different methods. In this experiment, we selected the CCDC29 gene as the amplification target and designed specific primers and probes for it to ensure accurate and efficient PCR amplification. Standard DNA samples were prepared by serial dilution from Quantifiler Duo DNA Standard, and the conversion efficiency was calculated using the formula, $T/(T+C) \times 100\%$. Meanwhile, multiplex ARMS-PCR products were analyzed using an ABI 3500 Genetic Analyzer, with a detection threshold of 50 rfu. Methylation values for each CpG were calculated using the formula T/(T+C), based on the relative peak area.

3. Results and discussion

3.1. Comparison of methylation values

Methylation values of 16 body fluid-specific CpGs were analyzed across 100 samples using both enzymatic and bisulfite conversion. Enzymatic conversion

demonstrated more pronounced differences in methylation values across body fluids and detected more outliers, particularly in blood-specific CpGs, where some samples resembled those of menstrual blood. Semen-specific CpGs showed concentrated distributions, indicating strong specificity. For blood-specific CpGs such as cg01607849 and cg06379435, methylation values were similar between blood and menstrual blood, especially after enzymatic conversion. In contrast, bisulfite conversion produced more concentrated values. For cg08792630 and cg26285698, enzymatic conversion resulted in wider distributions, while bisulfite conversion was more uniform. Saliva-specific CpGs generally maintained high specificity, though cg05761971 displayed a broader range. Vaginal secretion-specific CpGs (cg26285089 and cg25416153) showed overlapping values with non-target fluids, particularly in enzymatic conversion. Lastly, menstrual blood-specific CpG cg09696411 showed low methylation values across all fluids. The results are shown in Figure 1. Overall, enzymatic conversion provided greater variability in detecting methylation patterns.

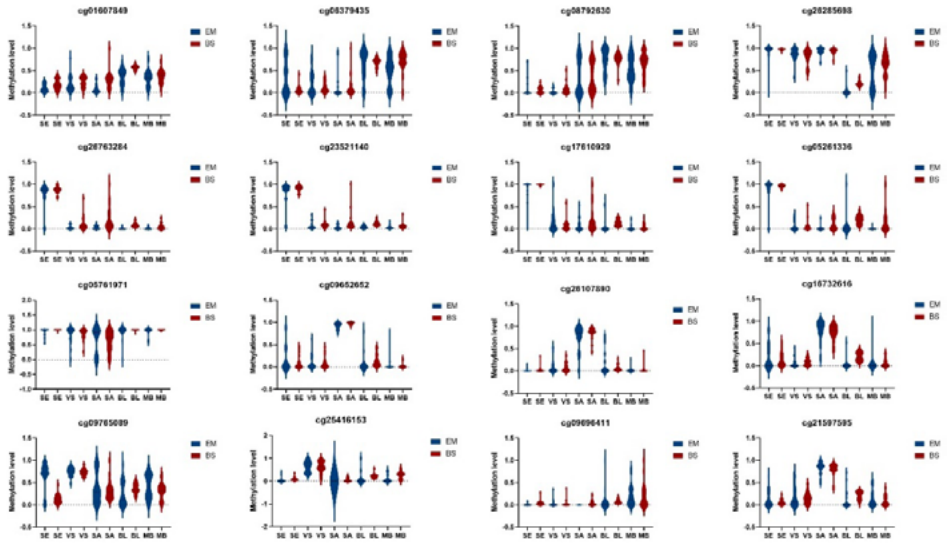


Fig. 1. Methylation values of 16 body fluid-specific CpGs analyzed across 100 samples using both enzymatic (EM) and bisulfite (BS) conversion methods. The X-axis represents different body fluids, including semen (SE), vaginal secretions (VS), saliva (SA), blood (BL), menstrual blood (MB), while the Y-axis represents methylation levels.

3.2. Comparison of conversion efficiency

In terms of conversion efficiency, enzymatic conversion demonstrated slightly lower efficiency compared to the bisulfite conversion. The average conversion efficiency and standard deviation for enzymatic conversion across 100 single body fluid samples was 0.841±0.079, while for bisulfite conversion, they were 0.881±0.056. The results are shown in Figure 2.

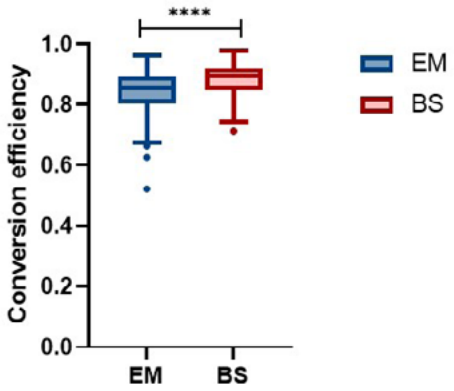


Fig. 2. Comparison of conversion efficiency between enzymatic (EM) and bisulfite (BS) methods. The X-axis represents the different conversion methods (EM and BS), while the Y-axis represents conversion efficiency. The results show a statistically significant difference between the two methods, ****$p < 0.0001$.

3.3. The limit of detection

To evaluate the detection limits of both conversion methods, a semen sample was diluted to DNA inputs of 150ng, 100ng, 50ng, and 10ng, with the conversions, amplification, and electrophoresis processes repeated three times. Figure 3 illustrates the mean methylation values obtained from these varying inputs. At DNA inputs ≥ 50 ng, both methods produced full methylation profiles, showing minimal differences between them. However, as DNA input decreased to 10ng, random CpG drop-out was observed in both methods. We noted that overestimation of methylation values when DNA input was low (≤10ng), especially in bisulfite conversion. These results suggest that the enzymatic conversion method is better suited for low-template samples.

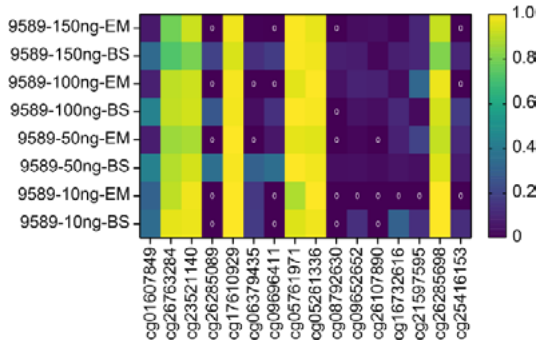


Fig. 3. Heatmap of mean methylation values across varying DNA inputs (150ng, 100ng, 50ng, and 10ng) for both enzymatic (EM) and bisulfite (BS) conversion methods. The X-axis represents different CpGs, while the Y-axis represents different DNA inputs.

4. Conclusion

This study explored the application of enzymatic conversion for DNA methylation analysis in forensic BFI, and compared it with bisulfite conversion. Although the enzymatic conversion method demonstrated slightly lower conversion efficiency compared to bisulfite conversion, the enzymatic method provided more stable results at low DNA input. Overall, the enzymatic conversion method shows considerable promise for BFI in forensic contexts and may be applicable in cases involving challenging samples.

5. Acknowledgement

The authors would like to thank sample donors participating in the study. Yihang Huang performed the experiments and drafted the manuscript. Yuanyuan Xiao conducted the statistical models and performed the experiments and data analysis. Manrui Li and Haoyan Jiang contributed to the sample collection and experiments. Yuxuan Tan contributed to the development and optimization of the system. Xiameng Chen and Shengqiu Qu contributed to the proofreading of the manuscript. Weibo Liang and Peng Bai conceived and supervised the project and designed the study. All the authors have read and approved the final version of the manuscript.

Conflict of interest statement

The authors declare no conflict of interest.

References

[1] Vaisvila R, Ponnaluri V K C, Sun Z, et al. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA[J]. Genome Research, 2021, 31(7): 1280-1289. doi.org/10.1101/gr.266551.120.

[2] Li Y, XU H, Chitaman J M, et al. Whole Genome DNA Methylation Profiling of D2 Medium Spiny Neurons in Mouse Nucleus Accumbens Using Two Independent Library Preparation Methods[J]. Genes, 2022, 13(2). doi.org/10.3390/genes13020306.

[3] Han Y, Zheleznyakova G Y, Marincevic-Zuniga Y, et al. Comparison of EM-seq and PBAT methylome library methods for low-input DNA[J]. Epigenetics, 2021, 17(10): 1195-1204. doi.org/10.1080/15592294.2021.1997406.

[4] Tian H, Bai P, Tan Y, et al. A new method to detect methylation profiles for forensic body fluid identification combining ARMS-PCR technique and random forest model[J]. Forensic Sci Int Genet, 2020, 49: 102371. doi.org/10.1016/j.fsigen.2020.102371.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 809-817
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Preliminary Exploration of tsRNAs in Forensic Body Fluid Source Identification

Jiaru Li[1], Hongwu Ren[1], Li Wang[2], Jing Zhu[3], Qian Wang[4,*], Zhilong Li[5,*]

[1]*West China School of Basic Medical Sciences and Forensic Medicine, Sichuan University, Chengdu, P.R. China;*

[2]*Department of Medical Genetics, Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University), West China Second University Hospital, Sichuan University, Chengdu, P.R. China;*

[3]*Forensic Science and Technology Department Sichuan Police College Luzhou, P.R. China;*

[4]*Laboratory of Metabolomics and Gynecological Disease Research, Key Laboratory of Birth Defects and Related Diseases of Women and Children of Ministry of Education, West China Second University Hospital, Sichuan University, Chengdu, P.R. China;*

[5]*Laboratory of Molecular Translational Medicine, Center for Translational Medicine, Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University), Ministry of Education, West China Second University Hospital, Sichuan University, Chengdu, P.R. China.*

** Corresponding authors at: No.20, People's South Road, Section 3, Wuhou District, Chengdu, China.*
wangqian151@foxmail.com (Q.Wang), lizhilong-cn@foxmail.com (Z.Li)

Abstract

A novel class of small RNAs, tRNA-derived small RNAs (tsRNAs) has been detected in various body fluid samples. However, the potential application of tsRNAs in forensic body fluid source identification has not yet been reported. In this study, we attempted to screen tsRNAs to distinguish between saliva and vaginal secretions, two body fluids originating from epithelial cells. Stem-loop reverse transcription followed by SYBR Green quantitative polymerase chain reaction (qPCR) was employed to measure the abundance of tsRNA, with U6 snRNA serving as the reference control. Following a series of experiments to ensure detection accuracy and Cq value adjustments, 10 tsRNAs were found to be differentially expressed between saliva and vaginal secretions. To further investigate the potential of tsRNA in forensic body fluid source identification, tsRNAs were detected in semen, peripheral blood, and menstrual blood. The results showed that 18 tsRNAs had significant differential expression across the five body fluids. By combining quantitative data with a logistic regression machine (MLR) learning model or neural networks (NN)

model, high accuracy in leave-one-out cross-validation (LOOCV) was achieved. Additional body fluid-specific tsRNAs are required and may enhance the accuracy of the prediction model while reducing the number of loci needed. Our study validated the potential of tsRNAs for forensic body fluid identification.

Keywords
Forensic Science, Body Fluid Source Identification, tsRNA, RT-qPCR, Machine Learning.

1. Introduction

Earlier, we developed a panel of 12 target microRNAs (miRNAs)s to differentiate between five body fluid types[1]. Using machine learning algorithms, the predictive models achieved 100% accuracy. Despite this strong performance, the model lacked vaginal secretion-specific miRNAs, presenting challenges for model simplification.

A novel class of noncoding RNAs, tRNA-derived small RNAs (tsRNAs), ranging from 14 to 40 nt in length, are abundant, highly modified, dynamically regulated small RNAs that interact with other RNAs or proteins[2]. As key regulators[3], tsRNAs are generated non-randomly through nuclease activity on tRNAs[4]. Growing evidence suggests that tsRNAs hold promise as biomarkers[5-7], and they have been detected in various body fluids[8]. Godoy et al.[8] further reported that tsRNAs exhibit distinct expression patterns across different body fluids, in contrast to miRNAs, indicating their potential for body fluid origin identification in forensic contexts.

To date, no studies have reported the use of tsRNAs for forensic body fluid identification. In the present study, we analyzed three saliva samples and three vaginal secretion samples using Arraystar small RNA microarray, which integrates both tsRNAs and miRNAs in a single array, allowing simultaneous detection of these two non-coding RNAs. The array results revealed that tsRNAs were more abundant than miRNAs, with a greater number of differentially expressed tsRNAs. Additionally, we developed a tsRNA detection method tailored for forensic applications and validated its accuracy through a series of experiments. Our study represents a novel attempt to apply tsRNAs in forensic body fluid source identification.

2. Material and methods

2.1. Sample collection

A total of five types of body fluids (saliva, vaginal secretion, peripheral blood, menstrual blood, and semen) were collected, with 15 samples for each type. All samples were obtained from self-reported healthy, unrelated individuals, with ethical approval from the Ethics Committee of West China Second University Hospital, Sichuan University. Written informed consent was provided by all donors, which included 32 males and 43 females, aged 19 to 46 years. This study adhered to the ethical guidelines outlined in the World Medical Association's Declaration of Helsinki[9]. The details of sample collection followed the same procedures as described in our previous publication[1].

2.2. RNA isolation and quality control

For array detection, total RNA was isolated from approximately 2 mL of saliva and a complete vaginal secretion swab using TRIzol (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The samples used for RNA extraction included about 2 mL of liquid saliva, 0.5 mL of liquid peripheral blood, 0.5 mL of liquid semen, a full vaginal secretion swab, and approximately 2 cm^2 of menstrual blood-soaked gauze. The gauze was cut into small pieces and dissolved in 0.5 mL of RNase-free water, while the vaginal secretion swab was immersed in 0.5 mL of RNase-free water for elution. Total RNA was then extracted from the resulting solutions.

2.3. Array analysis and candidates selection

The Array Small RNA Microarray (Arraystar, Rockville, Maryland, USA) includes over 14,000 probes designed to specifically quantify small RNA biotypes such as miRNAs, tsRNAs, pre-miRNAs, tRNAs, and snoRNAs. In brief, for each sample, 100 ng of total RNA was dephosphorylated to create a uniform 3'-OH end. The RNA was then denatured using DMSO and enzymatically labeled with Cy3. The Cy3-labeled small RNAs were hybridized to the Arraystar Human Small RNA Arrays in an Agilent Hybridization Oven. After hybridization, the slides were washed and scanned using an Agilent Scanner G2505C. Data processing and analysis were subsequently performed. Microarray scanning and data analysis were provided by KangChen Bio-tech. Candidate selection was primarily based on the array results for tsRNAs, which were ranked by fold change (FC) from high to low, with attention given to statistical significance (p-value) and signal intensity on the chip, as sites

with high FC but low signal values were difficult to detect. U6-snRNA was used as the reference gene.

2.4. RT-qPCR

We explored various methods to detect tsRNA expression and ultimately decided to use stem-looped specific reverse transcription with SYBR Green qPCR. The reverse primers were specifically designed according to the sequence of target tsRNAs. The length of tsRNAs were extended after reverse transcription facilitating subsequent quantitative detection. Quantitative primers were designed based on the cDNA sequences and universe part of reverse primers. To ensure the accuracy of detection, we amplified the products using traditional PCR, and checked the length of products by gel electrophoresis, then conducted TA cloning for sequencing. Total RNA with a series of dilution ratios (40ng to 0.039ng) were detected to evaluate amplification efficiency of the method. All experiments were conducted in strict accordance with the MIQE guidelines[10], ensuring the reliability and clarity of qPCR result interpretation.

2.5. Statistical analysis

Due to variations in amplification efficiency, all tsRNA Cq values were normalized to U6 snRNA, whose amplification efficiency was very close to two (corrected $Cq_{\text{Target tsRNA}}$= ordinary $Cq_{\text{Target tsRNA}}$*. The relative expression of tsRNA was calculated using the delta (ΔCq) method: ΔCq = corrected $Cq_{\text{Target tsRNA}} - Cq_{\text{U6-snRNA}}$. Normal distribution check was conducted using SPSS 26.0. Differences in tsRNA expression between saliva and vaginal secretions were analyzed using an independent samples t-test or a non-parametric test. Expression differences of tsRNA across the five body fluid types were evaluated using ANOVA or non-parametric tests. We applied Principal Component Analysis (PCA) to reduce the dimensionality of the dataset and visualize the distribution of different body fluid types based on tsRNA expression.

2.6. Development and refinement of the predictive model

Prediction model development and debugging *was implemented in Python, utilizing libraries such as pandas, scikit-learn, seaborn, and matplotlib for data handling, preprocessing, and visualization. Two machine learning methods, MLR and NN, were used to build the body fluid predictive model.*

3. Results

3.1.Quality control of samples and array experiments

All samples used in array screening experiment passed the quality control. Samples used in subsequent RT- qPCR was extracted successfully. The values of A260/A280 ranged from 1.80 to 2.12, and the concentration of samples ranged from 12 ng/uL (a saliva sample) to 528 ng/uL (a vaginal secretion sample). With the thresholds $|FC| \geq 1.5$ and $p < 0.05$, differentially expressed miRNAs and tsRNAs were screened. For miRNAs, 715 miRNAs were vaginal secretion-specific and 66 miRNAs were saliva-specific. For tsRNAs, 796 tsRNAs were vaginal secretion-specific and 720 tsRNAs were saliva-specific. Array signals indicated the abundance of tsRNAs was higher than miRNAs. Array results confirmed tsRNAs might hold body fluid-specific expression patterns which made it could be used in forensic body fluid identification.

3.2. tsRNA candidates selection

We selected 10 saliva-specific tsRNAs and 10 vaginal secretion-specific tsRNAs to perform subsequent quantification validation. Length of tsRNA candidates ranged from 16 nt (i-tRF-38:53-Cys-GCA-110 and tRF5-16-LeuAAG-1) to 44 nt (mt-3'tiRNA-44-LysTTT). U6 snRNA was selected as the reference gene to normalize the expression levels of the target tsRNAs.

3.3. Evaluating the accuracy of target tsRNAs and U6-snRNA

Specific stem-looped reverse transcription was used to generate cDNA, then cDNA was quantitatively detected using SYBR Green. All the sites yielded smooth amplification curves. The amplification curves of the sites were first checked, and tsRNA sites with double peaks were discarded (tRF3-17-TyrGTA-4 was discarded). And mt-3'tiRNA-41-HisGTG was excluded because the amplification fragment length did not align with the expected result. The remaining 18 tsRNAs and U6 snRNA were then subjected to TA cloning and sequencing to verify the accuracy of their sequences. The sequencing results confirmed that the sequences of these sites were correct.

3.4. Body fluid specificity of tsRNA candidates

The expression of all 18 tsRNAs was detected using RT-qPCR and analyzed across 10 saliva samples and 10 vaginal secretion samples. The expression of five tsRNAs (tRF5-18-LeuCAA-6, tRF5-17-LeuCAG-1, tRF5-17-LeuTAG-1, mt-tRF3b-

AlaTGC and mt-Trf3-28-LysTTT) was higher in saliva, while the abundance of another five tsRNAs (i-tRF-18:35-Gly-GCC-1, i-tRF-16:33-Thr-CGT-6, i-tRF-23:40-Gln-CTG-1, mt-tRF3-26-MetCAT and i-tRF-38:53-Cys-GCA-110) was greater in vaginal secretions. These ten tsRNAs had potential to distinguish between saliva and vaginal secretions.

3.5. The performance of tsRNA to identify five body fluids

We wonder how these 18 tsRNAs were expressed in three other common body fluids (menstrual blood, peripheral blood and semen), hence we detected an additional 30 samples using RT-qPCR. Quantitative results showed significant differences in the expression of the 18 tsRNAs across the five bodily fluids. Subsequently, we used PCA to analyze the 18 tsRNAs in the 50 samples. The PCA plot showed that the first two principal components explained 84.61% of the variance.

3.6. Assessment of the tsRNA-based machine learning prediction model

A total of 66 samples of tsRNA data were obtained from the above experiments, with 50 samples (10 from each of the five types of body fluids) used as the training set, and the remaining 16 samples as the test set. Two machine learning methods MLR and a NN classifier, were used to construct the prediction model.

3.6.1. MLR model

The model built using multiple logistic regression, utilizing the relative expression data of 18 tsRNAs, achieved a LOOCV accuracy of 94% on the 50 training samples. *Misclassifications included one vaginal secretion sample incorrectly identified as menstrual blood, one saliva sample misclassified as semen, and one menstrual blood sample misidentified as peripheral blood. Simultaneously, we utilized three methods to evaluate the importance of parameters from different perspectives: Random Forest Classifier, Recursive Feature Elimination (RFE), and SHAP (SHapley Additive exPlanations). Different methods yield varying results in the evaluation of locus importance. For the Random Forest Classifier, the top three loci contributing the most were i-tRF-18:35-Gly-GCC-1, tRF5-16-LeuAAG-1, and mt-tRF3-26-MetCAT. In contrast, the top three contributing loci identified by RFE were i-tRF-18:35-Gly-GCC-1, tRF5-16-LeuAAG-1, and i-tRF-23:40-Gln-CTG-1. Meanwhile, SHAP analysis highlighted mt-tRF3-26-MetCAT, mt-3'tiRNA-38-TrpTCA, and i-tRF-38:53-Cys-GCA-110 as the top contributors. The trained MLR model predicted the test set samples with an accuracy of 75%.*

3.6.2. NN model

Similarly, using the NN method to model the relative expression data of 18 tsRNAs from the training set, the confusion matrix shows a LOOCV accuracy of 96% (Figure 1A). Two vaginal secretion samples were misclassified as peripheral blood and menstrual blood, respectively. The permutation importance method was used to evaluate the importance of parameters, and the results show that the three loci with the greatest contribution were i-tRF-23:40-Gln-CTG-1, tRF3-21-ThrAGT-5, and i-tRF-38:53-Cys-GCA-110(Figure 1B). Using the full-parameter NN model to predict the test set, the prediction accuracy was 81.25%. One saliva sample and two vaginal secretion samples were misclassified as menstrual blood samples.

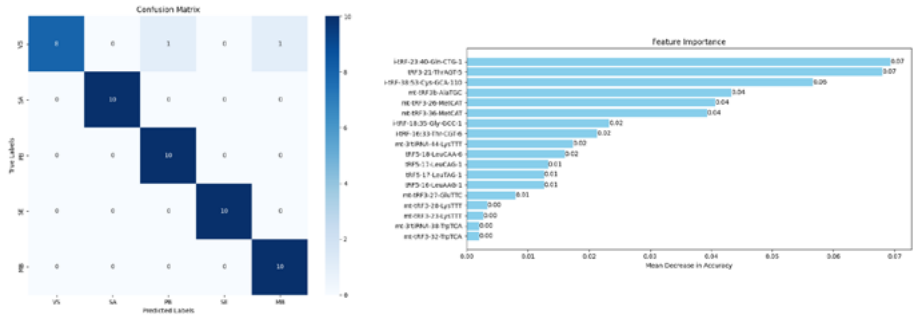


Figure 1. Performance and Feature Importance Evaluation of the NN Model. A, Confusion matrix plot for the training set of the NN model (VS, SA, PB, SE, MB refer to vaginal secretion, saliva, peripheral blood, semen and menstrual blood respectively). B, parameter importance evaluated using permutation importance method.

4. Discussion and Conclusion

Utilizing various modeling techniques to assess feature importance is crucial for validating their effectiveness and reducing the risk of overfitting, a common issue with LOOCV. Additionally, precise identification of highly specific markers is vital, and comparing them with miRNAs plays a key role in determining their potential as forensic biomarkers.

In this study, we explored the use of tsRNA to differentiate the origins of common body fluid stains. Candidate tsRNAs were identified using arrays and subsequently validated through RT-qPCR. A range of methods was implemented to

ensure the precision of detection targets. Our findings indicated that tsRNA could reliably distinguish between saliva and vaginal secretions—two body fluids that are rich in epithelial cells and are typically difficult to differentiate. The 18 tsRNAs identified showed distinct expression patterns across five common body fluids, highlighting the potential of tsRNA for forensic identification of body fluid samples.

5. Funding

This paper was supported by Sichuan Science and Technology Program (No. MZGC20230066).

6. Acknowledgments

None.

7. Conflict of interest statement

The authors declare no conflict of interest.

8. References

[1] Z. Li, M. Lv, D. Peng *et al.*, "Feasibility of using probabilistic methods to analyse microRNA quantitative data in forensically relevant body fluids: a proof-of-principle study," Int J Legal Med, vol. 135, no. 6, pp. 2247-2261, 2021. doi.org/10.1007/s00414-021-02678-w.

[2] J. Park, S.H. Ahn, M.G. Shin *et al.*, "tRNA-Derived Small RNAs: Novel Epigenetic Regulators," Cancers (Basel), vol. 12, no. 10, 2020. doi.org/10.3390/cancers12102773.

[3] C. Cole, A. Sobala, C. Lu *et al.*, "Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs," RNA, vol. 15, no. 12, pp. 2147-2160, 2009. doi.org/10.1261/rna.1738409.

[4] D. Haussecker, Y. Huang, A. Lau *et al.*, "Human tRNA-derived small RNAs in the global regulation of RNA silencing," RNA, vol. 16, no. 4, pp. 673-695, 2010. doi.org/10.1261/rna.2000810.

[5] J.M. Dhahbi, S.R. Spindler, H. Atamna *et al.*, "Deep Sequencing of Serum Small RNAs Identifies Patterns of 5' tRNA Half and YRNA Fragment Expression Associated with Breast Cancer," Biomark Cancer, vol. 6, pp. 37-47, 2014. doi.org/10.4137/BIC.S20764.

[6] V. Oberbauer and M.R. Schaefer, "tRNA-Derived Small RNAs: Biogenesis, Modification, Function and Potential Impact on Human Disease Development," Genes (Basel), vol. 9, no. 12, 2018. doi.org/10.3390/genes9120607.

[7] L. Zhu, J. Li, Y. Gong *et al.*, "Exosomal tRNA-derived small RNA as a promising biomarker for cancer diagnosis," Mol Cancer, vol. 18, no. 1, pp. 74, 2019. doi.org/10.1186/s12943-019-1000-8.

[8] P.M. Godoy, N.R. Bhakta, A.J. Barczak *et al.*, "Large Differences in Small RNA Composition Between Human Biofluids," Cell Rep, vol. 25, no. 5, pp. 1346-1358, 2018. doi.org/10.1016/j.celrep.2018.10.014.

[9] W.M. Association, "World Medical Association Declaration of Helsinki Ethical Principles for Medical Research Involving Human Subjects," JAMA, vol. 310, pp. 2191-2194, 2013. doi.org/10.1001/jama.2013.281053.

[10] C.T. Wittwer, J. Vandesompele, G.L. Shipley *et al.*, "The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments," Clinical Chemistry, vol. 55, no. 4, pp. 611-622, 2009. doi.org/10.1373/clinchem.2008.112797.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 819-827
DOI: https://dx.doi.org/10.15304/cc.2025.1869

What do forensic geneticists ignore? Two alternative, non-toxic solvents for FFPE tissue deparaffinization

Giulia Soldati[1], Dario Raniero[1], Chiara Saccardo[1], Francesco Ausania[1], Domenico De Leo[1], Stefania Turrina[1]

[1]*Department of Diagnostics and Public Health, Unit of Forensic Medicine, Forensic Genetics Lab, University of Verona, Italy*

Corresponding Author: *Giulia Soldati - E-mail: giulia.soldati_02@univr.it*
Forensic Genetics Laboratory – Institute of Legal Medicine - Department of Diagnostic and Public Health
University of Verona – Italy

Abstract

In forensic genetics, it is sometimes necessary to employ biological tissue fixed in formalin and subsequently embedded in paraffin (FFPE) as a source of DNA.
In the FFPE tissue pre-treatment step, deparaffinization plays a crucial role as it can significantly influence the outcome of the subsequent DNA extraction step.
Currently, the standard dewaxing procedure involves using xylene to remove paraffin from FFPE tissues. As it is an organic solvent highly toxic to the operator, different deparaffinization protocols using non-toxic solvents, such as white mineral oil and limonene, were compared in this study.
For this purpose, one FFPE lung tissue taken after the autopsy of a male body was used. The sample was cut into twenty slices by a microtome to allow the dewaxing protocols to be performed (five protocols, each repeated four times). DNA extraction from the tissue samples was performed using the QIAmp DNA Mini kit. The extracted DNA was quantified and then subjected to amplification with the PowerPlex® Fusion 6C system, after which amplicons were genotyped with the SeqStudio™ Genetic Analyzer for HID.
Based on the results of this preliminary study, it is apparent that the degree of informativeness of the genetic profiles varied depending on the deparaffinization technique employed: mineral oil demonstrated superior efficiency, yielding an average of 18 accurately typed markers out of 27.
While further validation on a more extensive and diverse set of FFPE tissues is necessary, it appears justifiable to posit that alternative solvents, possessing equivalent or higher efficacy but reduced or absent toxicity levels, could serve as viable substitutes for xylene in deparaffinization.

Keywords
Formalin Fixed Paraffin Embedded tissues, Deparaffinization, White Mineral Oil, DNA extraction

Introduction

Formalin fixation and paraffin embedding are a cornerstone in clinical diagnostics and biomedical research. Creating formalin-fixed paraffin-embedded (FFPE) tissue blocks is integral to histological analyses, providing pathologists with critical tools for conclusive diagnostics and research investigations. These FFPE blocks enable detailed cellular examinations through histochemical and immunohistochemical staining, offering enhanced visibility of cells of interest [1, 2].

One of the significant advantages of FFPE blocks is their long-term stability at room temperature, which makes them a cost-effective alternative to the labour-intensive and expensive storage of frozen tissues at ultra-low temperatures. As a result, FFPE tissue blocks are extensively archived, serving as a vast repository of biological material for ongoing and future research and diagnostics [1, 3-4]

Despite their widespread use and utility, FFPE tissues pose significant challenges when utilized for genetic and epigenetic analyses [1]. While preserving tissue integrity and allowing for detailed morphological and immunohistochemical assessments, the formalin fixation process introduces substantial difficulties in nucleic acid extraction. Formaldehyde, the fixative used, forms extensive cross-links between proteins and between proteins and nucleic acids, complicating the retrieval of high-quality DNA necessary for molecular analyses such as Polymerase Chain Reaction (PCR). Moreover, unbuffered formalin solutions can lead to acid-driven hydrolytic nucleic acid fragmentation, reducing extractable DNA yield and quality [3-7].

The standardization of FFPE sample processing varies among laboratories and is influenced by specific protocols, types of fixatives, fixation durations, and other reagents used. The penetration rate of formalin, approximately 1 mm per hour, necessitates careful control of fixation time to prevent excessive cross-linking and nucleic acid degradation. Consequently, optimizing fixation conditions is crucial to minimizing DNA fragmentation and retrieving sufficiently pure and intact nucleic acids for subsequent analyses [8].

In summary, the factors that, singularly or in combination, could pose significant difficulties in extracting DNA from FFPE are the type of fixative, the inclusion medium used, the method of preservation, the time between fixation/inclusion and DNA analysis, and the biological tissue itself [9-11].

Four steps are necessary to recover DNA from FFPE tissue sections: deparaffinization, tissue lysis, reverse-crosslinking, and DNA purification [1]. Since the pre-treatment process of the biological sample can significantly influence and negatively impact the subsequent DNA extraction step, tissue deparaffinization plays a crucial role.

Several protocols have already been developed and validated to extract high amounts of amplifiable DNA from FFPE tissues. Still, as it is difficult to establish with certainty which of the above-mentioned factors is predominant in influencing the quality and quantity of the extracted DNA, the development of alternative methods that consider the health of the operators in addition to the goodness of the results can be of support to the forensic geneticist community.

Therefore, this study compared the standard deparaffinization protocol with xylene to four other methods with non-toxic solvents, such as white mineral oil (WMO) and D-limonene, in which paraffin is soluble.

Materials and methods

To have a reference sample of the victim, which was less susceptible to chemical changes than the FFPE blocks, the DNA taken from the blood of the femoral artery during the autopsy was extracted using the QIAamp DNA Mini Kit.

Then, two FFPE blocks containing a lung tissue portion of the unknown male body were sectioned in twenty slices by a microtome to allow the deparaffinization protocols to be performed. Each of the five protocols was repeated four times. The first block was labelled with the code 5183-16, while the second one with the code 5183-17.

Informed consent was unnecessary in this study because the samples were derived from a judicial–autopsy case. The report does not contain personal information. All data are covered by the Italian Law – Data Protection Authority (Official Gazette no. 72 of March 26, 2012) for scientific research purposes.

Standard deparaffinization protocol – xylene (1)

Sectioned slices were immersed in 1 ml of xylene and vortexed for 10 min, followed by centrifugation at 12.000 rpm for 2 min. The supernatant containing the solubilized paraffin was discarded. This procedure was repeated twice with xylene and twice more, replacing xylene with 100% ethanol. Then, 1 ml of 75% ethanol was added to the pellet. After vortexing for 5 min, centrifuging at 12.000 rpm for 2

min, and discarding the supernatant, the same procedure was performed by adding 1 ml of 50% ethanol. Finally, leaving the cap of the 2 ml tube open, an incubation at 37°C for 10 min was applied.

Non-toxic deparaffinization protocols – white mineral oil (2)

Two deparaffinization protocols were tested: the former was tested within the laboratory for the first time, and the second one was a readjustment of a protocol proposed by Beckman Coulter [12]:

a. sectioned slices were immersed in 250 µl of WMO, vortexed, and incubated at 56°C for 10 min. The WMO was removed, and the procedure was repeated again.
b. sectioned slices were immersed in 450 µl of WMO, vortexed, and incubated at 80°C for 5 min. 200 µl of ATL was added, and the sample was centrifuged at 1000 rpm for 15 s. A second incubation at 80°C for 5 min was performed, after which the sample was allowed to cool down at room temperature for 2 min. Without discarding the WMO, 30 µl of Proteinase K was added, and the tissue was lysed for 2 hours at 56°C. The pellet was removed and transferred to a new 2 ml tube.

Non-toxic deparaffinization protocols – d-limonene (3)

Two deparaffinization protocols were tested: the former was the same standard protocol with xylene, in which the d-limonene replaced the latter, whereas the second one was the customized Beckman Coulter protocol, in which the d-limonene replaced the WMO.

Common DNA extraction and quantification

Regardless of the deparaffinization method, DNA was extracted using the QIAamp DNA Mini Kit, strictly adhering to the manufacturer's protocol. Specifically, for samples not deparaffinized using Beckman's protocol, 400 µl of ATL buffer was added to the deparaffinized sample, followed by vortexing for 2 min. Subsequently, 20 µl of Proteinase K was introduced, and the mixture was incubated at 56°C for 2 hours. For all samples, including tissues deparaffinized according to Beckman's protocol, 200 µl of AL buffer was added, with an additional incubation at 56°C for 10 min. Next, 200 µl of absolute ethanol was added, and the resulting mixture was applied to a spin column provided with the kit. The column was centrifuged at 8,000 rpm for 1 min, washed with 500 µl of AW1 buffer, and then with

500 µl of AW2 buffer. DNA elution was performed using 45 µl of AE buffer, following a 3-min incubation at room temperature.

The DNA quantification was performed with the Qubit Fluorimeter using the Qubit dsDNA HS Assay Kit and samples were normalized to 1 ng/µl.

DNA amplification, and typing

The genomic DNA of all samples was quantified and then amplified with the PowerPlex® Fusion 6C System (referred to as Fusion 6C), following the manufacturer's indications. Amplicons were genotyped with the SeqStudio™ Genetic Analyzer for HID, and fragment analysis was conducted with GeneMapper ID-X v1.6 Software.

Results and discussion

The DNA concentrations obtained after applying the different deparaffinization protocols and the subsequent DNA extraction are shown in Table 1(A).

DNA concentrations were reproducible depending on the block used and the deparaffinization method applied.

The DNA typing performed on blood extracted from the femoral artery produced a complete profile of 24 aSTRs and 3 Y-STRs. This profile was used to verify the accuracy of allelic designation in profiles obtained from DNA extracted from lung tissue sections of the same unknown male subject.

None of the 20 typing results of the tissue sections produced a complete genetic profile. Specifically, the genetic profiles with the highest number of correctly typed markers were obtained from sections 2a.1 (19 aSTRs and 3 Y-STRs) and 2b.2 (20 aSTRs and 2 Y-STRs), both of which underwent deparaffinization with WMO. The genetic profile with the lowest number of correctly typed markers was found in section 3b.1 (9 aSTRs and 1 Y-STR).

On average, tissue sections deparaffinized with d-limonene produced the least satisfactory genetic profiles, with an average of 12 correctly typed markers. The results are summarized in Table 1(B).

Table 1 A) DNA quantification results; B) DNA typing result: the number of corrected, incomplete, and not genotyped loci are reported, in addition to the average of corrected genotyped markers

				A	B			
Deparaffinization solvent	Protocol	Repetition	FFPE block ID	DNA quantification [ng/µl]	N. of corrected genotyped loci	N. of incomplete genotyped loci	N. of not genotyped loci	Average of corrected genotyped loci
XYLENE	1	1	5183-17	7,6	16	4	7	13
		2	5183-16	0,18	14	2	11	
		3	5183-16	0,7	11	2	14	
		4	5183-16	0,7	12	1	14	
WHITE MINERAL OIL	2a	1	5183-17	12,6	22	3	2	17
		2	5183-17	9,5	18	3	6	
		3	5183-16	1,4	13	1	13	
		4	5183-16	2,3	13	4	10	
	2b	1	5183-17	45	20	2	5	19
		2	5183-17	45,3	22	1	4	
		3	5183-16	2,72	18	3	6	
		4	5183-16	5,7	17	2	8	
d-LIMONENE	3a	1	5183-17	0,7	14	2	11	12
		2	5183-17	0,7	12	1	14	
		3	5183-16	0,3	11	1	15	
		4	5183-16	0,7	11	1	15	
	3b	1	5183-17	2,15	10	5	12	12
		2	5183-17	4	12	2	13	
		3	5183-16	3,2	12	4	11	
		4	5183-16	3,2	14	3	10	

This study compared five deparaffinization protocols, starting with the solvent currently used in the standard deparaffinization procedure adopted by most forensic genetics and pathology laboratories, namely xylene. Since it is a highly toxic solvent for operators, as evidenced by the Globally Harmonized System of Classification and Labelling of Chemicals, it was necessary to evaluate deparaffinization protocols that excluded the use of harmful organic solvents. Based on the chemical principle of *similia similibus solvuntur*, WMO and d-limonene were chosen.

The individual genetic profile analysis showed a higher or lower loss of sample informativeness depending on the deparaffinization method used. The best efficiency was obtained by exploiting WMO, which allowed correct genotyping of up to 22 of the 27 analyzed loci. This finding confirms that using WMO as a dewaxing solvent is non-toxic to the operator and results in highly superior genetic profiles compared to other solvents.

Conclusions

FFPE tissue blocks are invaluable for long-term storage and various diagnostic and research applications. Still, the challenges associated with DNA extraction from

these samples highlight the need for refined protocols and methods to maximize their utility in forensic studies.

This study used three solvents to dewax paraffin from FFPE lung tissues: xylene, WMO, and d-limonene. The best results regarding the number of corrected genotyped loci were obtained using WMO, which is non-toxic for the operator.

Although this is a preliminary work that needs validation on a more significant number of different FFPE biopsy samples and tissues, it seems more than reasonable to assume that, as an alternative to the use of xylene, deparaffination can be performed with other equally effective solvents with less or no toxic characteristics. The aim of the work is also to urge forensic genetics laboratories to make a significant change in their laboratory practices, improving both the safety of operators and the quality of the results while remaining in line with the European Council of Legal Medicine (ECLM)-validated guidelines [13].

Acknowledgements

None

Conflict of interest statement

None

References

[1] Oba U, Kohashi K, Sangatsuda Y, Oda Y, Sonoda KH, Ohga S, Yoshimoto K, Arai Y, Yachida S, Shibata T, Ito T, Miura F. An efficient procedure for the recovery of DNA from formalin-fixed paraffin-embedded tissue sections. Biol Methods Protoc. 2022 Jul 26;7(1):bpac014. doi.org/10.1093/biomethods/bpac014. Erratum in: Biol Methods Protoc. 2023 Mar 13;8(1):bpad003. doi.org/10.1093/biomethods/bpad003. PMID: 35937639; PMCID: PMC9351614.

[2] Vitošević, K., Todorović, M., Slović, Ž., Varljen, T., Matić, S., & Todorović, D. (2021). DNA isolated from formalin-fixed paraffin-embedded healthy tissue after 30 years of storage can be used for forensic studies. Forensic science, medicine, and pathology, 17(1), 47–57. https://doi.org/10.1007/s12024-020-00327-z

[3] Lin, J., Kennedy, S. H., Svarovsky, T., Rogers, J., Kemnitz, J. W., Xu, A., & Zondervan, K. T. (2009). High-quality genomic DNA extraction from formalin-fixed and paraffin-embedded samples deparaffinized using mineral oil. Analytical biochemistry, 395(2), 265–267. https://doi.org/10.1016/j.ab.2009.08.016

[4] Schmeller, J., Wessolly, M., Mairinger, E., Borchert, S., Hager, T., Mairinger, T., Schmid, K. W., Wohlschlaeger, J., Walter, R. F. H., & Mairinger, F. D. (2019). Setting out the frame conditions for feasible use of FFPE derived RNA. Pathology, research and practice, 215(2), 381–386. https://doi.org/10.1016/j.prp.2018.12.027

[5] Paireder, S., Werner, B., Bailer, J., Werther, W., Schmid, E., Patzak, B., & Cichna-Markl, M. (2013). Comparison of protocols for DNA extraction from long-term preserved formalin fixed tissues. Analytical biochemistry, 439(2), 152–160. https://doi.org/10.1016/j.ab.2013.04.006

[6] Nguyen, H. T., Tatipamula, V. B., Do, D. N., Huynh, T. C., & Dang, M. K. (2022). Retrieving high-quality genomic DNA from formalin-fixed paraffin-embedded tissues for multiple molecular analyses. Preparative biochemistry & biotechnology, 52(1), 48–55. https://doi.org/10.1080/10826068.2021.1923030

[7] Talukdar, F. R., Abramović, I., Cuenin, C., Carreira, C., Gangane, N., Sincic, N., & Herceg, Z. (2022). A protocol for good quality genomic DNA isolation from formalin-fixed paraffin-embedded tissues without using commercial kits. Molecular biology reports, 49(5), 4115–4121. https://doi.org/10.1007/s11033-022-07394-1

[8] Funabashi, K. S., Barcelos, D., Visoná, I., e Silva, M. S., e Sousa, M. L., de Franco, M. F., & Iwamura, E. S. (2012). DNA extraction and molecular analysis of non-tumoral liver, spleen, and brain from autopsy samples: the effect of formalin fixation and paraffin embedding. Pathology, research and practice, 208(10), 584–591. https://doi.org/10.1016/j.prp.2012.07.001

[9] Greer, C. E., Peterson, S. L., Kiviat, N. B., & Manos, M. M. (1991). PCR amplification from paraffin-embedded tissues. Effects of fixative and fixation time. American journal of clinical pathology, 95(2), 117–124. https://doi.org/10.1093/ajcp/95.2.117

[10] Romero, R. L., Juston, A. C., Ballantyne, J., & Henry, B. E. (1997). The applicability of formalin-fixed and formalin fixed paraffin embedded tissues in forensic DNA analysis. Journal of forensic sciences, 42(4), 708–714.

[11] Benavides, J., García-Pariente, C., Gelmetti, D., Fuertes, M., Ferreras, M. C., García-Marín, J. F., & Pérez, V. (2006). Effects of fixative type and fixation time on the detection of Maedi Visna virus by PCR and immunohistochemistry in paraffin-embedded ovine lung samples. Journal of virological methods, 137(2), 317–324. https://doi.org/10.1016/j.jviromet.2006.07.009

[12] FormaPure XL DNA: Extended Protocol for DNA Isolation from FFPE Sample. Instructions for use. Beckman Coulter, PN C40291AB, April 2019.

[13] Mangin, P., Bonbled, F., Väli, M., Luna, A., Bajanowski, T., Hougen, H. P., Ludes, B., Ferrara, D., Cusack, D., Keller, E., & Vieira, N. (2015). European Council of Legal Medicine (ECLM) accreditation of forensic pathology services in Europe. International journal of legal medicine, 129(2), 395–403. https://doi.org/10.1007/s00414-014-1041-x

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 829-835
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Comparisons of DNA extraction methods for telogen hair samples

Cathrine Bie Petersen, Floor Claessens, Maryam Sharafi Farzad, Helle Smidt Mogensen, Claus Børsting, Marie-Louise Kampmann*

Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

**Corresponding author: marie.louise.kampmann@sund.ku.dk*

Abstract

Telogen hair samples are valuable sources of DNA for forensic genetics and commonly found during crime scene investigations. Traditionally, mitochondrial DNA (mtDNA) has been the most common target for forensic genetic analyses of telogen hair extracts. However, with the introduction of shotgun sequencing, small, highly degraded fragments of nuclear DNA (nuDNA) from hair may be used for human identification. Here, we tested the efficacy of three DNA extraction methods designed for hair, and we examined and compared yields of both mtDNA and nuDNA. The chosen extraction methods are commonly used for isolating mtDNA in forensic genetics and ancient DNA research.

Two fragments were analysed from each telogen hair: A 2 cm fragment including the root, and a 2 cm fragment proximate to the root fragment. The hair fragments were cleaned using an ultrasonic bath and a proteinase solution. Digestion was performed according to the published protocols. The digestion buffer contained varying amounts of Proteinase K and DTT, and the digestion time ranged from 30 minutes to overnight incubation. The extraction methods also varied in the purification step, where the techniques for isolation of DNA were based on either glass fibres or silica attached to magnetic particles or in a column.

The results confirmed that both mtDNA and nuDNA are present in hair samples. The preliminary results indicated that higher DNA yields are obtained using an extraction method with glass fibre DNA isolation. Overall, our study provides valuable insights into DNA extraction methods for telogen hair samples and highlight the importance of method selection.

Keywords

Telogen hair, hair extraction, quantification, mitochondrial DNA, nuclear DNA

Introduction

Telogen hairs are commonly encountered at crime scenes and could serve as critical evidence in forensic investigations. However, the highly degraded state of DNA in telogen hairs presents significant challenges for forensic genetic analyses [1]. As a result, mitochondrial DNA (mtDNA), which is more abundant and stable than nuclear DNA (nuDNA), has traditionally been the preferred target for forensic genetic testing of hair samples [2]. However, with advancements in sequencing technologies, particularly shotgun sequencing, it is now possible to analyse even smaller and more degraded fragments of both nuDNA and mtDNA, offering new opportunities for genetic analysis of compromised samples.

This study evaluates the efficacy of three different DNA extraction methods for telogen hairs, concerning both nuDNA and mtDNA yield. The first method (Method A) is adapted from ancient DNA research [3], a field that frequently investigates degraded samples. The second method (Method B) is an in-house protocol originally designed for extracting DNA from anagen hairs [4]. The third method (Method C) was previously developed for the extraction of DNA from telogen hairs [1].

To assess the performance of these methods, we extracted DNA from both root end (n=60) and rootless (n=60) fragments of telogen hairs obtained from ten unrelated donors. The extracted DNA was quantified using three different qPCR assays, targeting both nuDNA and mtDNA fragments of varying sizes to determine the mtDNA copy number and the nuDNA quantity [5-7]. The SD quants [5] quantifies a 70 bp nuDNA target together with mtDNA targets of 69 bp (short fragment) and 143 bp (long fragment). The Triplex [6] amplifies two mtDNA targets of 105 bp (short fragment) and 316 bp (long fragment). Finally, Quantifiler™ Trio (Thermo Fisher Scientific) quantifies nuDNA targets of 80 bp (short fragment) and 214 bp (long fragment). The ratio between short and long fragments for each of the quantification methods produces a degradation index (DI) [7].

This comparative analysis provides insights into the efficacy of different extraction techniques for telogen hair samples, with potential implications for improving forensic genetic analyses of hair.

Table 1. Overview of principal elements and conditions of the compared extraction methods

	Method A ancient DNA method	Method B in-house method	Method C telogen hair method
Digestion buffer	Tris-HCl, NaCl, $CaCl_2$, SDS, DTT, Proteinase K, EDTA, phenol Red	Tris-HCl, NaCl, $CaCl_2$, SDS, DTT, Proteinase K	ATL (Qiagen), DTT, Proteinase K
Binding buffer	Buffer PB (Qiagen), NaOAc, NaCl, phenol Red	EZ1&2® Investigator® Reagent Cartridge (Qiagen)	Isopropanol
Extraction	Manual	Automated	Manual
DNA binding	Glass	Silica	Silica
Immobilisation	Column Viral Nucleic Acid Large Volume (Roche)	Magnets EZ1&2® Investigator® Reagent Cartridge (Qiagen)	Magnets Prepfiler™ (Thermo Fisher Scientific)
Elution	TE	Water	TE
Reference	[3]	[4]	[1]

Materials & methods

Telogen hairs were donated from ten individuals and every individual donated six hairs. The telogen state was confirmed by microscopic inspection and hairs were sectioned into two fragments: a 2 cm fragment including the root (root-end), and the adjacent 2 cm fragment (rootless).

Fragments from two hairs from each donor were extracted in duplicate using each of the extraction methods listed in Table 1. Prior to extraction all hair fragments underwent a Terg-a-zyme wash as described in Brandhagen et al. [1] and subsequently, extractions were performed as described in [1, 3, 4]. DNA was eluted in 50 µL buffer or water. Quantifications were performed as described in [5, 6] and the Quantifiler™ Trio User Guide. Statistical tests and plots were made using R version 4.3.1.

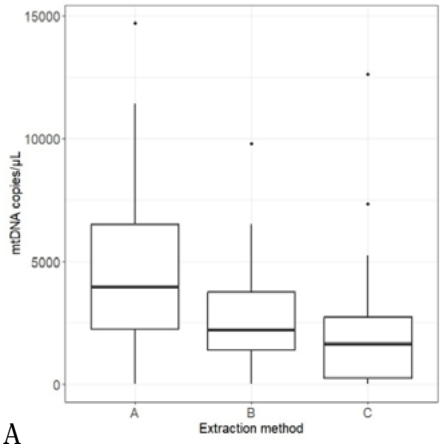
Results

The efficacy of the three different extraction methods (Methods A-C) was evaluated using hair shafts from the same cohort of ten individuals. Mitochondrial DNA

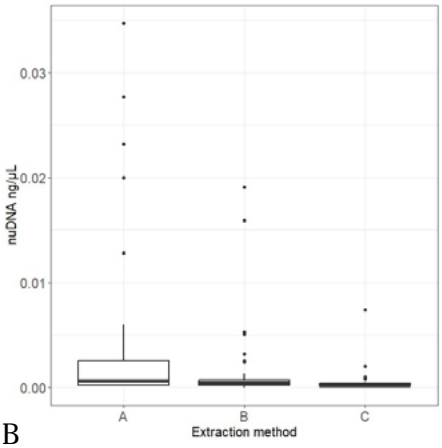
(mtDNA) quantification from both root-end and rootless hair fragments demonstrated that extraction Method A was the most efficient (Figure 1A). Method B yielded intermediate mtDNA concentrations, while Method C produced the lowest concentrations. Statistical analysis using ANOVA revealed that mtDNA yields from Method A were significantly higher compared to both Method B ($p = 0.00137$) and Method C ($p = 0.0001$). Additionally, nuDNA quantities assessed via an 80 bp fragment from Quantifiler™ Trio and a 70 bp target in SD quants (4) revealed median nuDNA concentrations of 1.3, 0.7, and 0.6 pg/µL (Quantifiler™ Trio), and 1.8, 0.6, and 0.8 pg/µL (SD quants) for Methods A, B, and C, respectively (SD quants results, Figure 1B).

The root-end hair fragment resulted in up to several fold larger yield of both mtDNA and nuDNA compared to the adjacent root-less section as represented by the yield ratios between root-end and rootless end in Figure 1C&D.

The number of mtDNA copies and the nuDNA concentrations were determined using two qPCR assays: SD quants and Triplex. No amplification was observed for the 316 bp target in the Triplex assay, indicating that the nucleotide fragments in hair are very short. However, mtDNA copy numbers determined from both of the SD quants mtDNA targets showed a strong correlation with those obtained from the 105 bp target in the Triplex assay (Pearson coefficient R^2=0.92 and R^2=0.91). The comparison of mtDNA copy numbers between the two SD quants targets and the 105 bp Triplex target suggests that the Triplex assay may underestimate mtDNA copy numbers relative to the SD quants assay (Figure 1E).



A



B

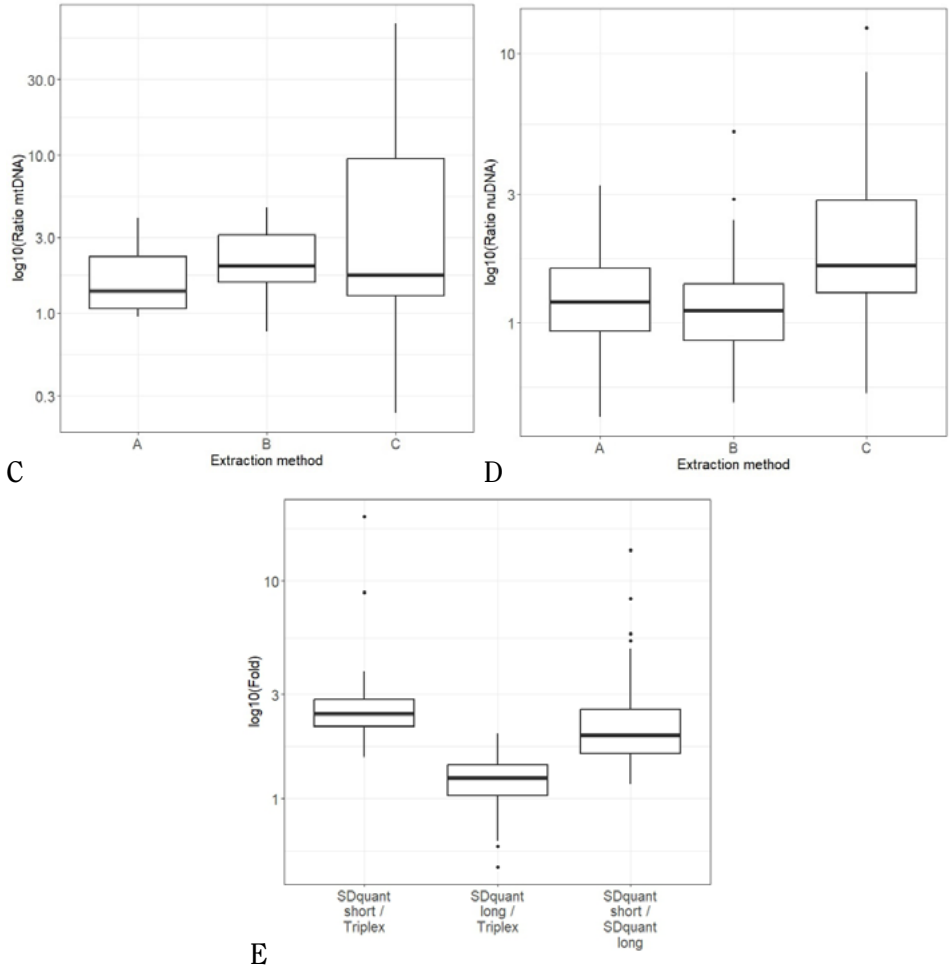


Figure 1. A) Comparison of the yield obtained from the three extraction methods based on the short mitochondrial DNA fragment (69 bp) from SD quants quantification. **B)** Comparison of the yield obtained from the three extraction methods based on the nuclear DNA fragment (70 bp) from SD quants quantification. **C)** The ratio of concentrations from root-end and the adjacent rootless hair fragment based on the short mitochondrial DNA fragment (69 bp) from SD quants. **D)** The ratio of concentrations from root-end and the adjacent rootless hair fragment based on the nuclear DNA fragment (70 bp) from SD quants. **E)** The log10 ratio of the concentrations from SD quants short fragment, Triplex short fragment and SD quants long fragment. The ratio of SD quants short to SD quants long can be interpretated as a Degradation Index for SD quants quantifications.

Discussion

In forensic investigations, the collection of hair fragments from crime scenes is expected to contribute to case resolution by providing valuable biological traces. When analysing hair shafts, obtaining a reliable short tandem repeat (STR) profile is often challenging due to the difficulty in amplifying long DNA fragments. However, mtDNA profiling can still be achieved and may offer useful insights. Since mtDNA is maternally inherited and not unique to individuals, nuDNA profiling, particularly using single nucleotide polymorphisms (SNPs), is essential for providing a unique genetic signature. Therefore, the selection of an optimal extraction method for both mtDNA and nuDNA is crucial to enhance the effectiveness of forensic investigations.

Our study demonstrated that the highest DNA yields were obtained using extraction Method A, which was originally developed for the recovery of DNA from ancient protein-rich sample material. This method likely captures numerous small DNA fragments, resulting in higher overall yields. The root-end of telogen hairs consistently produced the highest DNA concentrations. However, rootless hair should not be disregarded, as our data indicated that it may also yield high-quality DNA for forensic analyses. Based on these findings, we recommend including the root-end in extractions when available, but the entire hair shaft remains a viable source of DNA.

In this study, we used three qPCR assays targeting various fragment lengths to compare quantification methods. The selection of an appropriate assay should consider the target DNA fragment length required for downstream analyses. The Triplex assay, which targets a relatively long 316 bp fragment, was not suitable for determining hair-derived mtDNA as this fragment was not amplified in any of the hairs and no degradation index (DI) could be determined. In contrast, the SD quants assay, which utilises two shorter targets, enabled mtDNA quantification and estimation of DI value and is thus more appropriate for hair DNA analysis.

Conclusion

Telogen hair contains both mtDNA and nuDNA, which hold significant potential for forensic genetic analyses. Due to the fragmented nature of DNA in hair samples, the use of conventional standard short tandem repeats (STRs) analysis is often challenging. However, our findings demonstrated the successful recovery of high copy numbers of mtDNA, alongside substantial amounts of nuDNA, providing a valuable resource for forensic applications despite the degraded state of the DNA.

Extraction Method A yielded significantly higher concentrations of both mtDNA and nuDNA compared to Methods B and C. While the root-end of telogen hair provided the highest DNA yields, rootless hair fragments frequently contained sufficient DNA for shotgun sequencing analyses. The mtDNA qPCR assays produced comparable results, and the choice of assay should be guided by the size of the DNA fragments required for subsequent analyses.

Acknowledgements

The authors would like to thank Anja Ladegaard Jørgensen, Liselotte Bjerregård-Börner, Radoslaw Dawid Szymanski, Jette Pultz and Rikke Rune Larsen for laboratory assistance.

Conflict of interest statement

The authors declare no conflicting interest.

References

[1] Brandhagen MD, Loreille O, Irwin JA. Fragmented Nuclear DNA is the Predominant Genetic Material in Human Hair Shafts. Genes (Basel). 2018;9(12).

[2] Syndercombe Court D. Mitochondrial DNA in forensic use. Emerg Top Life Sci. 2021;5(3):415-26.

[3] Pedersen MW, Antunes C, De Cahsan B, Moreno-Mayar JV, Sikora M, Vinner L, et al. Ancient Human Genomes and Environmental DNA from the Cement Attaching 2,000-Year-Old Head Lice Nits. Mol Biol Evol. 2022;39(2).

[4] Hellmann A, Rohleder U, Schmitter H, Wittig M. STR typing of human telogen hairs--a new approach. Int J Legal Med. 2001;114(4-5):269-73.

[5] Xavier C, Eduardoff M, Strobl C, Parson W. SD quants-Sensitive detection tetraplex-system for nuclear and mitochondrial DNA quantification and degradation inference. Forensic Sci Int Genet. 2019;42:39-44.

[6] Kavlick MF. Development of a triplex mtDNA qPCR assay to assess quantification, degradation, inhibition, and amplification target copy numbers. Mitochondrion. 2019;46:41-50.

[7] ThermoFisher Scientific. Quantifiler™ HP and Trio DNA Quantification Kits USER GUIDE.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 837-842
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Degradation of DNA Extracts Stored Under Different Conditions: What We Know and What Is New. Preliminary Results

Sarah Gino[1*], Giulia Sguazzi[2], Selena Cisana[3], Luciano Garofano[4], Paolo Garofano[3], Pasquale Linarello[5], Monica Omedei[3], Luca Salvaderi[5], Noemi Procopio[6]

[1]*Department of Health Sciences, University of Eastern Piedmont, Novara, Italy & Direzione Medica dei Presidi Ospedalieri, AOU Maggiore della Carità, Novara, Italy*

[2]*CRIMEDIM, Center for Research and Training in Disaster Medicine, Humanitarian Aid and Global Health, University of Eastern Piedmont, Novara, Italy.*

[3]*Centro Regionale Antidoping "A. Bertinaria," Orbassano, Italy*

[4]*Italian Academy of Forensic Science, Parma, Italy*

[5]*Eurofins Genoma Group, Milano, Italy*

[6]*Research Centre for Field Archaeology and Forensic Taphonomy, School of Law and Policing, University of Central Lancashire, Preston, UK;*

** Corresponding author at: Department of Health Sciences, University of Eastern Piedmont, Via Solaroli 17, 28100 Novara, Italy, sarah.gino@uniupo.it, +39 0321660644*

Abstract

Preserving the integrity of forensic evidence through the stabilisation of the biological signature contained in DNA extracts is particularly important, specifically when re-analysing samples after a certain period of time. This translates into the needs for the appropriate storage of DNA extracts to ensure the successful outcome of forensic genetic analyses, including human identification through STR typing and the application of phenotype and ancestry sequencing panels to infer the morphological characteristics and the geographical origin of the contributor. Although guidelines shared by the international scientific community recommend freezing of the DNA extracts for long-term storage, it is possible that samples are preserved in sub-optimal conditions for a variety of reasons (e.g., faulty freezers, moving samples to court offices at the case closure). In these circumstances, little is known about the survival of DNA and its suitability for future STR and SNP analyses.

In this study, we evaluated the preservation of DNA extracts obtained from buccal swabs collected from two subjects of different sexes using the QIAamp DNA Investigator Kit (Qiagen). The extracts were quantified (Quantifiler™ Trio DNA Quantification Kit – Thermo Fisher Scientific) and aliquots containing 1 ng/µl of DNA were created. In addition, mixtures were set up from the

female and male DNA extracts, in a ratio of 1:15 (male/female). All the aliquots were stored at three temperatures (+20°C, +4°C and -20°C) for a period of 150 days. As the results at 90 days showed a surprising DNA survival rate also at room temperature, revealing similar completeness of the STR profiles (GlobalFiler™ PCR Amplification Kit – Thermo Fisher Scientific) at both -20°C and +20°C conditions, the experiment was further extended to 400 days. In addition, after the initial 90 days experiment, a subset of samples was moved to uncontrolled temperature conditions. Surprisingly, DNA extracts did not degrade up to 400 days under any condition, and the ratio in the mixed samples remained unchanged. It was also noted that storing the extracts at +4°C and +20°C caused their evaporation and consequently increased their concentration up to 20-folds. Overall, these results could open new avenues for the analysis of samples from cold cases preserved in sub-optimal conditions.

Keywords
Human identification, DNA storage, DNA degradation.

1. Introduction

Preserving the integrity of forensic evidence through the stabilization of the biological signature contained in DNA extracts is crucial, especially when their re-analysis is required after a certain period of time [1]. This translates into the needs to appropriately preserve DNA extracts to ensure the successful outcome of forensic genetic analyses, including human identification through STR typing and the application of phenotypic and ancestry panels to infer the morphological characteristics and geographical origin of the contributor [2]. Although the guidelines recommend freezing for long-term storage, samples may be preserved under sub-optimal conditions due to various reasons (e.g., faulty freezers, relocation to court offices) [3-5]. Under these circumstances, little is known about the survival of the DNA and its suitability for future STR and SNP analyses.

2. Material studied, methods, techniques

Twelve and twenty-four buccal swabs were collected, respectively, from a male (black hair, black eyes, Indian) and a female (blonde hair, blue eyes, Italian). They were extracted using QIAcube and QIAamp Investigator kits. The extracts were quantified using Quantifiler™ Trio DNA Quantification Kit (Thermo Fisher Scientific), diluted and aliquoted in 5 replicates each, resulting in a DNA concentration of 1 ng/µl and a volume of 20 µl. Mixtures were also prepared from the female and male DNA extracts, in a 15:1 ratio (female/male).

Aliquots were stored under three different conditions (-20°C, +4°C and +20°C) for 150 days, with regular assessments of degradation (at 0-7-15-30-90-150 days). After 90 days, all replicates that had already been analysed and stored at -20°C, +4°C and +20°C were divided into two subgroups that were transferred to room controlled temperature and to room not controlled temperature, respectively, breaking the cold chain. These samples were quantified after 310 days.

Samples kept at the starting condition for 150 days after quantification were preserved at the same condition for another 250 days and quantified again.

It is important to note that the total storage time was 400 days for all samples analysed, even if at different conditions.

Replicates (male, female and mixture) with the highest degradation index or abnormal concentration values under all storage conditions were amplified using the GlobalFiler PCR Amplification Kit and sequenced with SeqStudio Genetic Analyzer. GeneMapper® ID-X v1.5 software was used to define the genetic profiles. This study belongs to a larger project submitted and approved by the Novara Intercompany Ethics Committee (CE 24/21).

3. Results

Quantifications performed at 0,7,15,30 and 90 days showed a degradation index below or equal to 1 for all three storage conditions (+20°C, +4°C and -20°C). Based on the 90-days results, which showed that DNA concentration and mixture ratios remained relatively constant at all three temperatures, the cold chain was interrupted, and measurements were taken again after an extended period (310 days after the interruption). In contrast, after 400 days of preservation at room temperature, the degradation index increased to values greater than 1. Based on the quantifications and degradation index, non-standard storage temperatures (+4°C, +20°C and room temperature) do not significantly affect DNA quality but can impact DNA quantity through evaporation (observed in 73% of cases). Breaking the cold chain resulted in having a degradation index greater than 1 only when samples were transferred from -20°C to uncontrolled storage conditions. Samples moved and stored at +20°C and uncontrolled temperature experienced a 15-fold increase in DNA concentration due to evaporation. Evaporation and concentration (a 20-fold increase in quantification) also affected samples stored at +20°C for 400 days. The success and failure rates, along with the challenges encountered due to suboptimal conservation conditions, are summarized in Figure 1.

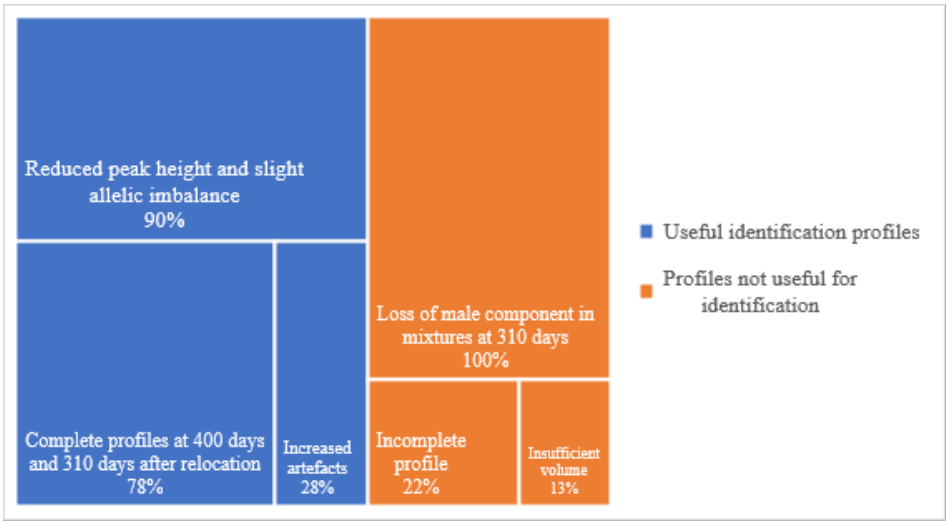


Figure 1. Tree map showing the percentages of useful profiles for personal identification and non-useful profiles for identification.

Evaporation of the samples led sometimes to insufficient volume for amplification and typing of the genetic profile in some cases. As an examples, one of the replicates of the male extracts stored originally at +20°C for 90 days and then moved to an uncontrolled temperature for 310 days, had a DNA concentration of 27.3 ng/µl and a degradation index of 1.30. Due to insufficient volume, direct dilution in the tube was necessary before adding a suitable volume to the amplification reaction. Despite this, a useful genetic profile for identification was obtained (24/24 typed loci), which did not show any artefacts (drop-in/drop-out).

4. Discussion

Surprisingly, our preliminary results suggest that DNA extracts did not degrade even after 400 days under any conditions, and the male-female ratio in mixed samples remained unchanged throughout the study. Storage at unusual temperatures and interruptions of the cold chain can affect STR typing, particularly in terms of peak height and balance, with the greater effects the greater the temperature deviation. Storing extracts at +4°C and +20°C/room temperature for prolonged times led to sample evaporation, with a consequent increase in their concentration

up to 20-folds. Evaporation and increased concentration also occurred following the interruption of the cold chain. Finally, the example of male profile illustrates that, despite an insufficient volume, the direct dilution in the tube can allow for a successful amplification and for the typing of a complete genetic profile.

5. Conclusion

In conclusion, unconventional storage procedures can effectively preserve DNA samples for extended periods of time, as demonstrated by the high DNA recovery and integrity rates. Interruptions of the cold chain have a more significant impact on DNA preservation than their constant maintenance at sub-optimal temperatures for prolonged periods. Overall, these preliminary results may open new possibilities for analysing samples from cold cases that may not have been stored frozen. Given the potential evaporation of the sample and the resulting effect on DNA concentration, this approach complements traditional preservation methods and may be advantageous in some situations.

6. Acknowledgements

This research is the result of a study conducted in the framework of the International PhD in Global Health, Humanitarian Aid, and Disaster Medicine organized by University of Eastern Piedmont (UPO).

7. Conflict of interest statement

All authors disclose any financial and personal relationships with other people or organizations that could inappropriately influence their work.

8. References

[1] T. Graversen, J. Mortera, G. Lago, The Yara Gambirasio case: Combining evidence in a complex DNA mixture case, Forensic Sci. Int. Genet. 40 (2019) 52–63. https://doi.org/10.1016/j.fsigen.2018.12.010.

[2] C. Frippiat, S. Zorbo, D. Leonard, A. Marcotte, M. Chaput, C. Aelbrecht, F. Noel, Evaluation of novel forensic DNA storage methodologies, Forensic Sci. Int. Genet. 5 (2011) 386–392. https://doi.org/10.1016/J.FSIGEN.2010.08.007.

[3] S. Smith, P.A. Morin, Optimal Storage Conditions for Highly Dilute DNA Samples: A Role for Trehalose as a Preserving Agent, J Forensic Sci. 50 (2005). www.astm.org (accessed October 3, 2022).

[4] S.E. Howlett, H.S. Castillo, L.J. Gioeni, J.M. Robertson, J. Donfack, Evaluation of DNAstableTM for DNA storage at ambient temperature, Forensic Sci. Int. Genet. 8 (2014) 170–178. https://doi.org/10.1016/j.fsigen.2013.09.003.

[5] C. Frippiat, F. Noel, Efficiency of a novel forensic room-temperature DNA storage medium, Forensic Sci. Int. Genet. 9 (2014) 81–84. https://doi.org/10.1016/j.fsigen.2013.11.009.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 843-849
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Effects of fingerprint enhancement techniques on forensic DNA recovery

Xiaoyang Li[1*], Yong Sheng Lee[1], Susan Yueru Chua[1], Ying Yin Loh[1], Shawn Siong Chun Foo[2], Alvin Kok Leong Khah[2], Sheryl Su Yun Seet[2], Nur Ayuni Binte Sulaiman[2], Christopher Kiu-Choong Syn[1]

[1]*DNA Profiling Laboratory, Applied Sciences Group, Health Sciences Authority, 11 Outram Road, Singapore 169078, Singapore.*

[2]*Forensics Division, Criminal Investigation Department, Singapore Police Force, 391 New Bridge Road, Singapore 088762, Singapore.*

** Corresponding author at: DNA Profiling Laboratory, Applied Sciences Group, Health Sciences Authority, 11 Outram Road, Singapore 169078 - Email: Li_Xiaoyang@hsa.gov.sg - Tel: +65 66842053*

Abstract

Latent fingerprints and touch DNA are common traces found at crime scenes and can often be deposited onto a surface simultaneously through a single touch. As the development of fingerprints and the recovery of DNA may interfere with each other, Crime Scene Specialists (CSSs) often have to choose one method which is likely to offer the best chance for identification. In this study, we have assessed the impact of various fingerprint enhancement techniques on DNA recovery. The effects of these techniques on trace DNA recovery were examined on commonly encountered items in casework, including cellophane tape, duct tape, paper, plastic bag, glass, and acrylic. We have compared the quantity of extracted DNA and the quality of obtained DNA profiles from treated fingerprint samples with the controls. Our results showed that the application of wet powder suspension on the adhesive side of cellophane tape and the application of 1,8-Diazafluoren-9-one (DFO) and ninhydrin on paper caused decreased DNA recovery from the respective items. In contrast, other examined fingerprint enhancement techniques did not significantly affect DNA recovery from the corresponding items. Our study has demonstrated that fingerprints developed by various techniques could be successfully utilized for DNA analysis. It will also provide guidance for CSSs to select fingerprint enhancement processes that are compatible with DNA recovery, if necessary.

Keywords

Fingerprint, Fingerprint enhancement, DNA recovery, Touch DNA, DNA transfer.

1. Introduction

Latent fingerprints are common traces found at crime scenes and typically form an essential part of the evidence. Analysis of these fingerprints is a crucial tool in criminal investigations, which often requires the application of enhancement techniques [1]. Meanwhile, advancements in forensic DNA analysis have improved the sensitivity of DNA recovery, enabling the possibility of obtaining interpretable DNA profiles from biological materials left behind with fingerprint deposits [2]. Therefore, conducting DNA analysis after enhancing latent fingerprints may provide additional leads, especially in cases when the analysed fingerprint did not provide sufficient information to aid the investigation. However, concerns exist about the potential interference of fingerprint enhancement processes on subsequent DNA analysis, including extraction, amplification, and profiling [3].

In this study, we have assessed the potential effect of seven different fingerprint enhancement techniques on downstream DNA recovery from six types of items. Both the quantity of recovered DNA and the quality of DNA profiles were compared between the controls and the treated samples. The result of this study can provide guidance to Crime Scene Specialists (CSSs) in deciding appropriate fingerprint enhancement techniques during the investigation.

2. Material studied, methods, techniques

2.1. Preparation of biological samples

Volunteers were asked to wash their hands thoroughly with soap and water, then dab dry with clean paper towels. After that, they rubbed their fingers against their own foreheads, then deposited fingerprints and touch DNA through pressing their fingers against one of the studied items. These items included cellophane tape, duct tape, paper, plastic bag, glass, and acrylic. The touched items were then treated with one of the seven fingerprint enhancement techniques (wet powder suspension, gentian violet, 1,8-Diazafluoren-9-one (DFO) and ninhydrin, cyanoacrylate fuming, cyanoacrylate fuming and fluorescent dye, regular powder, and magnetic powder). DNA deposited on the studied items was collected with the double (wet-dry) swabbing technique [4,5]. Untreated samples were used as controls.

2.2. DNA sample processing

Cotton swab heads were excised with a scalpel blade, added to individual microcentrifuge tubes and subjected to DNA analysis. DNA extraction was performed using the DNA IQ™ Casework Extraction Kit (Promega) for pre-processing and DNA IQ™ Casework Pro Kit (Promega) for extraction on Maxwell® FSC instrument (Promega). DNA yields were estimated using Quantifiler® Trio DNA Quantification kit (Applied Biosystems) on QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystems). Extracted DNA was amplified with GlobalFiler™ kit (Applied Biosystems) with maximum of 15 µL or 1 ng input DNA on ProFlex PCR System (Applied Biosystems). Capillary electrophoresis was performed on ABI PRISM® 3500xL Genetic Analyzer (Applied Biosystems) with maximum injection parameters of 3 µL at 1.2 kV 24 s. Results were analyzed using GeneMapper® ID-X v1.6 software according to laboratory guidelines.

2.3. Statistical analyses

Wilcoxon signed-rank test was used to determine differences in DNA recovery between the controls and the treated fingerprint samples.

3. Results

Six different types of items commonly encountered in casework were selected in this study, including cellophane tape, duct tape, paper, plastic bag, glass, and acrylic. Fingerprints and touch DNA were deposited onto the surface of each item followed by corresponding fingerprint enhancement treatment as described. *Deposited biological materials were collected using the double swab* technique *and subjected to DNA analysis. The quantity of recovered DNA was compared between the treated samples and the controls to evaluate the effect of corresponding fingerprint enhancement techniques on DNA recovery. As shown in Figure 1, the application of wet powder suspension on the adhesive side of cellophane tape and the application of DFO and ninhydrin on paper led to significantly decreased DNA recovery; and an average decrease of 88% and 76% in extracted DNA amount were observed, respectively.* In contrast, other examined fingerprint enhancement techniques did not significantly affect DNA recovery from the corresponding items, including the application of gentian violet on the adhesive side of duct tape, cyanoacrylate fuming (with or without fluorescent dye) on plastic bag, regular powder on glass, and magnetic powder on acrylic.

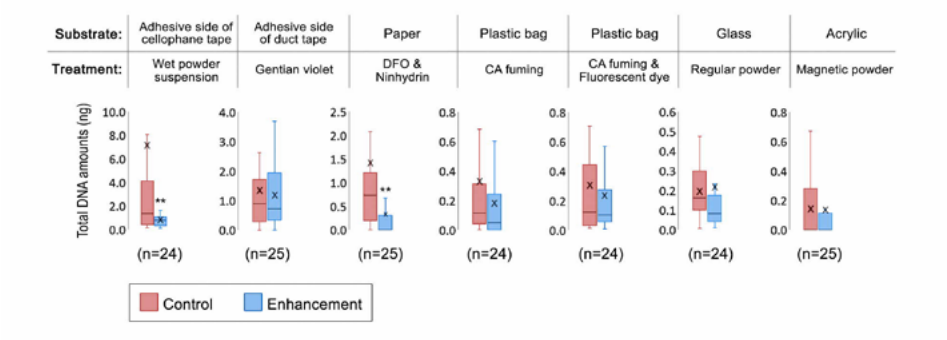


Figure 1. Quantity of DNA recovered from different types of items. Box and whisker plots with interquartile ranges, "x" represents the mean value. CA: cyanoacrylate. **p < 0.01

In addition, the quality of recovered DNA was also examined by comparing the percentage of recovered autosomal DNA alleles between the *controls and the treated samples* (Figure 2). *Application of DFO and ninhydrin on paper significantly reduced the number (average 59% decrease) of recovered autosomal DNA alleles after the fingerprint enhancement. In contrast, other examined fingerprint enhancement techniques did not affect the number of recovered autosomal DNA alleles significantly. After treating the adhesive side of cellophane tape with wet powder suspension, we observed a significant decrease in the amount of DNA recovered (Figure 1). However, this decrease in DNA yield did not have a comparably significant effect on the quality of the DNA profiles (Figure 2). This was attributed to the relatively higher quantity of DNA initially recovered from the cellophane tape as compared to paper,* owing to its adhesive characteristics.

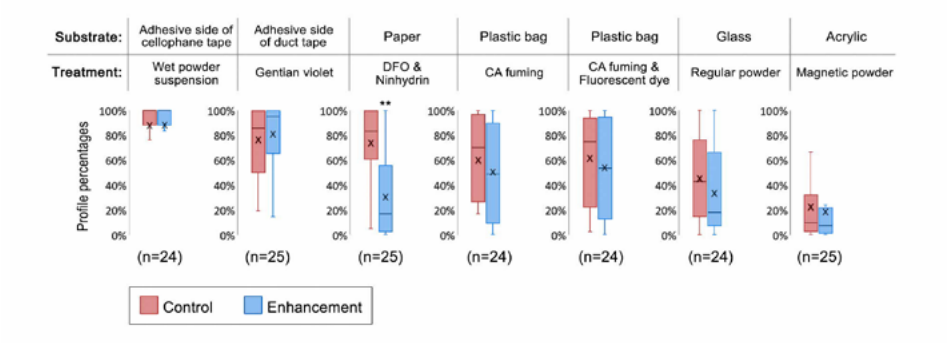


Figure 2. Percentage of recovered autosomal DNA alleles from different types of items. Box and whisker plots with interquartile ranges, "x" represents the mean value. A total of 21 autosomal DNA loci (GlobalFiler™ kit) were examined. Loci with a single allele were counted as 1 allele if the peak height was below laboratory stochastic threshold (RFU < 535), otherwise they were counted as 2 alleles. CA: cyanoacrylate. **p < 0.01

4. Discussion

The successful development of latent fingerprints is often critical in crime investigations. However, in cases where fingerprints fail to provide sufficient information, the biological materials left behind alongside with the prints can serve as a valuable source for developing DNA profiles to identify the person of interest (POI) [6]. Previous studies have shown that the swabbing process is usually destructive to fingerprints [7]. Therefore, visualization and recording of fingerprint details are typically performed prior to DNA analysis to maximize evidence recovery. The recoverability of DNA from developed latent fingerprints depends on both the type of items where fingerprints were deposited onto and the applied enhancement techniques [3]. Seven different combinations of item and enhancement process were tested in this study; and the results from three of them corroborated previous studies on the effect of DFO and ninhydrin, cyanoacrylate fuming, and regular powder on DNA recovery from paper, plastic bag, and glass, respectively [8,9]. Additionally, important and instructive results were also obtained here from the study of other combinations, including the application of wet powder suspension on the adhesive side of cellophane tape, gentian violet on the adhesive side of duct tape, cyanoacrylate fuming and fluorescent dye on plastic bag, and magnetic powder on acrylic. Our results here showed that relatively more DNA can be recovered from items with adhesive surfaces (such as cellophane tape and duct tape) or porous

surfaces (such as paper) compared to non-porous surfaces (such as plastic bag, glass, and acrylic). This is possibly due to the higher retention rate of biological materials on adhesive and porous surfaces.

Although DNA recovery can be significantly affected by the application of wet powder suspension on the adhesive side of cellophane tape or the application of DFO and ninhydrin on paper, interpretable or even full DNA profiles can still be generated from these post-fingerprint enhancement samples. In this study, an average of 37 and 13 autosomal DNA alleles were recovered from cellophane tape and paper after corresponding treatments, respectively, which could assist law enforcements to identify the POI.

Thus, in the situation where both latent fingerprints and touch DNA can be used as important evidence, CSSs should take a holistic view of the case and consider the crime context such as involved activities and types of surfaces, before deciding on which fingerprint enhancement technique to use.

5. Conclusion

DNA was successfully recovered from the fingerprints on different types of items after the application of fingerprint enhancement techniques, suggesting that fingerprints developed by various techniques can be successfully utilized for subsequent DNA analysis. Our results also indicated that both fingerprints and DNA profiles could be recovered from a single source when gentian violet, cyanoacrylate fuming (with or without fluorescent dye), regular powder, and magnetic powder were used on the adhesive side of duct tape, plastic bag, glass, and acrylic, respectively. However, CSSs may need a more considered approach when deciding to use wet powder suspension on the adhesive side of cellophane tape or DFO and ninhydrin on paper, as these treatments may have a detrimental effect on the subsequent DNA recovery.

6. Acknowledgments

This work was supported by the Health Sciences Authority of Singapore and Singapore Police Force. The authors would like to thank all volunteers for their contribution to this study.

7. Conflict of interest statement

The authors declare that they have no conflict of interest.

8. References

[1] B. Abebe, H.C.A. Murthy, E. Amare Zereffa, et al., Latent Fingerprint Enhancement Techniques: A Review, *J. Chem. Rev.* 2 (1) (2020) 40–56.

[2] H.R. Dash, K.M. Elkins, N.R. Al-Snan, Technological Advancements in DNA Extraction and Quantification of Forensic Samples. In: *Advancements in Forensic DNA Analysis*, Springer, Singapore, 2023, pp. 23–33.

[3] P. Kumar, R. Gupta, R. Singh, et al., Effects of latent fingerprint development reagents on subsequent forensic DNA typing: A review, *J. Forensic Legal Med.* 32 (2015) 64–69.

[4] B.C.M. Pang, B.K.K. Cheung, Double swab technique for collecting touched evidence, *Leg. Med.* 9 (4) (2007) 181–184.

[5] T.J. Verdon, R.J. Mitchell, R.A.H. van Oorschot, Swabs as DNA collection devices for sampling different biological materials from different substrates, *J. Forensic Sci.* 59 (4) (2014) 1080–1089.

[6] J. Burrill, B. Daniel, N. Frascione, A review of trace "Touch DNA" deposits: Variability factors and an exploration of cellular composition, *Forensic Sci. Int. Genet.* 39 (2019) 8–18.

[7] S. Fieldhouse, R. Parsons, S. Bleay, et al., The effect of DNA recovery on the subsequent quality of latent fingermarks: a pseudo-operational trial. *Forensic Sci Int.* 307 (2020) 110076.

[8] J. Sewell, I. Quinones, C. Ames, et al., Recovery of DNA and fingerprints from touched documents. *Forensic Sci. Int. Genet*. 2 (4) (2008) 281-285

[9] J.J. Raymond, C. Roux, E.D. Pasquier, et al., Effect of Common Fingerprint Detection Techniques on the DNA Typing of Fingerprints Deposited on Different Surfaces. *J. Forensic Identif.* 54 (1) (2004) 22-44

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 851-855
DOI: https://dx.doi.org/10.15304/cc.2025.1869

A 25 Loci Mini-STR System to Facilitate Analysis of Degraded DNA

Sheng Li[1]

[1]*NUHIGH Biotechnologies Co., Ltd. Suzhou, China.*

Abstract

DNA from crime scene evidence is often degraded, leading to incomplete profiles due to locus or allele dropout, which undermines DNA evidence reliability. Large fragments in forensic DNA kits are particularly prone to dropout. To address this, we developed the NHID® Mini25A Kit, a 25-loci mini-STR system designed for degraded samples, with amplicons under 275 bp. The kit shows excellent efficiency (PCR time: 54 minutes) and high performance in specificity, sensitivity, and inhibitor tolerance. It demonstrates robust sensitivity for trace DNA (50 pg) and high tolerance to five common inhibitors (500 mM hematin, 150 ng/μL humic acid, 10 mM indigo, 200 ng/μL collagen, and 1.5 mM CaCl2). Compared to the VeriFiler™ Plus PCR Kit, NHID® Mini25A shows a higher detection rate for degraded samples. Validated under SWGDAM guidelines, it has been tested in over 20 laboratories, confirming its forensic efficacy.

Keywords

Mini-STR, Degraded DNA, Forensic Science.

1. Introduction

Forensic DNA testing often faces degradation issues due to factors like temperature, humidity, UV exposure, and improper storage [1]. Degraded DNA complicates the amplification of long fragments, leading to incomplete profiles. Most STR kits amplify products between 70–550 bp, limiting their application in analyzing degraded samples like decomposed tissues, bones, and teeth, especially in old cases. Mini-STR technology addresses this by designing primers closer to STR core sequences, reducing amplicon lengths while preserving allelic information. Traditional mini-STR systems face limitations due to fluorescent dye channels and genetic analyzer capabilities. To overcome this, we developed the NHID® Mini25A Kit, a 25-loci mini-STR multiplex system, including all CODIS markers, Penta D, Penta

E, D6S1043, and Y-indel [2]. The kit uses two panels (A and B), with three shared loci for consistency verification (Figure 1).

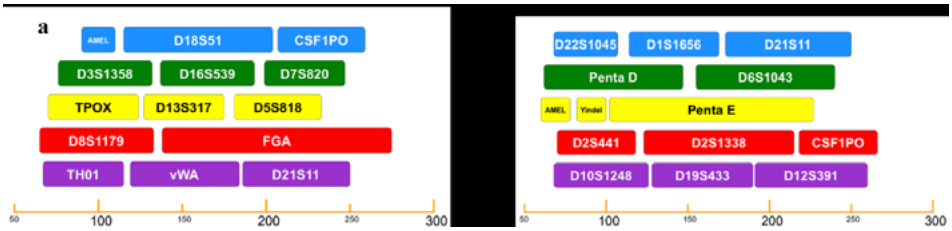


Figure 1. Marker positioning in the NHID® Mini25A Kit

2. Materials and methods

2.1. DNA sample preparation

Control DNA 9948 (0.5ng/μL) and genomic DNAs from non-primate animals (pig, cow, sheep, chicken, duck, rabbit, mouse, rat, fish, and horse) were sourced from NUHIGH Biotechnologies Co., Ltd. (Suzhou, China). DNA from human-associated microbes (e.g., *Escherichia coli*, *Lactobacillus delbrueckii*, *Candida albicans*, *Staphylococcus aureus*, *Streptococcus salivarius*, *Bacillus subtilis*, and *Neisseria gonorrhea*) [3] was obtained from Bena Culture Collection (Suzhou, China). A ten-year-old tooth sample was provided by a public security organization.

2.2. PCR amplification and thermal cycling conditions

The NHID® Mini25A Kit was optimized for 25 μL reactions, including 2.5 μL Primer Mix, 5 μL Master Mix, and up to 17.5 μL DNA. Cycling conditions: 95°C for 1 min (activation), 29 cycles of 95°C for 10s (denaturation), 58°C for 60s (annealing/extension), and final extension at 60°C for 5 min.

2.3. Sample electrophoresis and data analysis

Spectral calibration was performed with NH-6V Matrix before use with the 3500xL Genetic Analyzer. PCR products were mixed with formamide/size standard solution for capillary electrophoresis. Samples were injected at 1.2kV for 24 seconds and electrophoresed at 15 kV for 850 seconds. Data were analyzed using GeneMapper ID-X Software v1.6, with a peak threshold of 150 RFU.

2.4. Species specificity studies

To evaluate specificity, 1 ng of non-primate DNA and 20 ng of microbial DNA were amplified. Cross-reactivity was assessed using a 150 RFU threshold.

2.5. Sensitive studies

Serial dilutions of Control DNA 9948 were prepared with concentrations ranging from 500 to 31.25 pg per PCR reaction.

2.6. Inhibition studies

The NHID® Mini25A Kit was tested against five inhibitors: 500 µM hematin, 150 ng/µL humic acid, 10 mM indigo, 200 ng/µL collagen, and 1.5 mM CaCl2, using 0.25 ng Control DNA 9948.

2.7. Degraded DNA test

Degraded DNA from a ten-year-old tooth sample was tested with both NHID® Mini25A Kit and VeriFiler™ Plus PCR Kit. Allele detection rates (DR) and likelihood ratios (LR) were compared.

3. Results and discussion

3.1. Species specificity studies

The NHID® Mini25A Kit exhibited excellent specificity, with no cross-reactivity peaks above the 150 RFU threshold for non-primate or microbial DNA.

3.2. Sensitivity studies

Control DNA 9948 was detected in both panels at ≥50 pg, with consistent allele calls across four replicates (Figure 2).

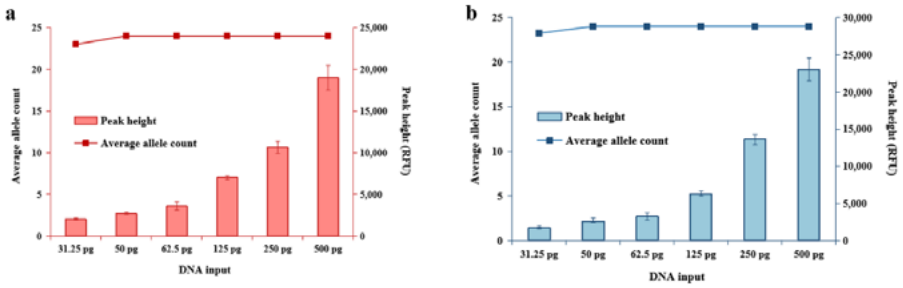


Figure 2. The average peak heights and number of alleles called at varying amounts of 9948 DNA.

3.3. Inhibition studies

Full STR profiles (42 alleles for Control DNA 9948) were obtained for all five inhibitors (Figure 3).

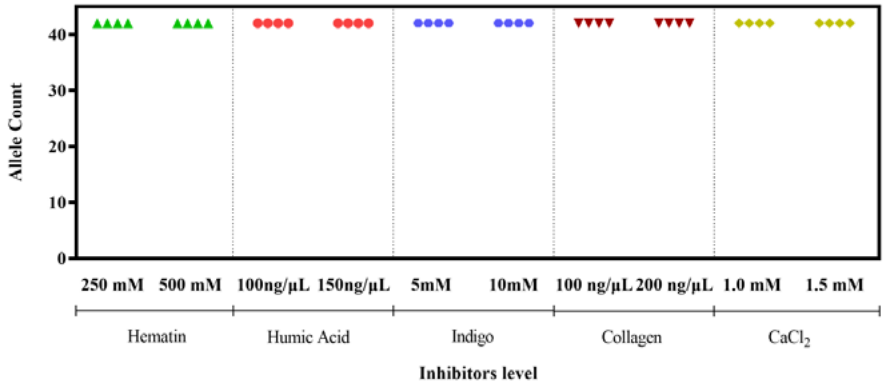


Figure 3. The tolerance of different inhibitors of NHID® Mini25A kit.

3.4. Degraded DNA test

In comparative testing, NHID® Mini25A outperformed the VeriFiler™ Plus Kit on degraded samples. For a ten-year-old tooth sample, 9 loci with amplicons >250 bp dropped out with VeriFiler™ Plus, while all 25 loci were successfully called with NHID® Mini25A. Detection rates for VeriFiler™ Plus and NHID® Mini25A were 64% and 100%, respectively (Table 1).

Table 1. The DR and LR of Verifiler™ Plus and NHID® Mini25A

	loci number detected	DR	LR
Verfiler Plus	16	64%	1.60E+15
NHID Mini25A	25	100%	5.31E+27

4. Conclusion

The NHID® Mini25A Kit is designed for degraded samples, with all amplified fragments under 275 bp. The kit exhibits excellent specificity, sensitivity, and inhibitor tolerance. It has been validated under SWGDAM guidelines and confirmed effective by multiple laboratories.

5. Acknowledgments

The authors would like to thank NUHIGH Biotechnologies Co., Ltd. and the Criminal Police Detachment of Yangzhou Public Security Bureau for their assistance and valuable feedback on this work.

6. Conflict of interest statement

The authors declare no conflicts of interest.

7. References

[1] Machida M, Kibayashi K. Effectiveness of whole genome amplification prior to short tandem repeat analysis for degraded DNA [J]. Forensic Science International: Genetics, 2020, 49: 102373.

[2] National Technical Committee for Standardization of Criminal Technology. Forensic sciences—data structures of selected loci from the DNA database: GB/T 41009-2021 [S]. 2021.

[3] Lee, S., McAnany, P.K., Chang, C.W. et al. Developmental validation of Applied Biosystems YFiler Platinum Casework PCR Amplification Kit. Sci Rep 13, 14539 (2023).

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 857-864
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Analysis of forensic samples using the Investigator Quantiplex Pro RGQ Kit (Qiagen) at half volumes

A. Barbaro*[123], A. LaMarca[4], G. Falcone[2]

[1]*Universidad de Alcalá, Departamento de Química Analítica, Química Física e Ingeniería Química, Ctra. Madrid-Barcelona km 33,6, 28871 Alcalá de Henares, Madrid, Spain*

[2] *Dept. Forensic Genetics,Studio Indagini Mediche E Forensi (SIMEF), via Nicolò da Reggio 4, 89128 Reggio Calabria, Italy*

[3]*Universidad de Alcalá, Instituto Universitario de Investigación en Ciencias Policiales, Libreros 27, 28801 Alcalá de Henares, Madrid, Spain.*

[4]*Dept. Forensic Medicine, Studio Indagini Mediche e Forensi (SIMEF), via Nicolò da Reggio 4, 89128 Reggio Calabria, Italy*

**Corresponding Author. Email: a.barbaro@uah.es; simef_dna@tiscali.it*

Abstract

The Investigator Quantiplex Pro RGQ Kit (Qiagen) employs fast-cycling technology to quantify total human and male DNA within approximately one hour using the Rotor-Gene Q instrument. It also provides insights into inhibitors, DNA degradation, and mixture conditions. This study evaluated the kit's performance with halved reaction volumes to enhance sample throughput per kit.

Dilution of Control DNA (50 ng/µl) and DNA extracted from 10 forensic casework samples, obtained following internal procedures and informed consent, were analyzed to check the performance of the Investigator Quantiplex Pro RGQ Kit, using half reaction volumes while maintaining component ratios. Results were analyzed using Q-Rex software, evaluating quantification data, inhibition (IC), degradation (DI) and mixture (MI) indices.

Standard curves for regular and halved volumes demonstrated efficient amplification of human, male, and degradation markers, with $R^2 > 0.998$ and appropriate slopes. Control samples showed strong linear correlation between theoretical and measured DNA concentrations for both conditions. Variability in DNA quantity and quality across forensic samples reflected substrate differences in preserving and releasing biological material. Nonetheless, concordant results were observed for each sample between halved and standard reaction volumes. This study confirms the Investigator Quantiplex Pro RGQ Kit delivers accurate and precise results with reduced reaction

volumes. This modification enables the analysis of more samples per kit, lowering laboratory costs without compromising data quality.

Key words
Quantiplex Pro RGQ Kit; DNA Quantification; Half Volumes; Forensic Samples

1. Introduction

DNA in forensic samples is often present in small quantities or degraded, leading to incomplete profiles. Challenges include PCR inhibition, DNA degradation, mixed profiles, and low DNA amounts. Inhibitors may arise from sample collection, preservation, substrates, or environmental exposure. [1-3] Quantification is a critical step in forensic DNA analysis, assessing DNA quantity and quality, particularly in degraded samples. [4]

Recent kits, like the Investigator Quantiplex Pro RGQ Kit (Qiagen), combine high sensitivity with robust inhibitor tolerance. Using fast-cycling technology, this kit quantifies DNA in approximately one hour with the Rotor-Gene Q Real-Time instrument. It employs four standards to generate a standard curve and targets five distinct regions: small and large human autosomal, small and large human male, and an internal control. This enables precise detection of total and male DNA, autosomal and Y-chromosomal degradation, and PCR inhibition monitoring.

The kit exhibits high sensitivity, detecting low male DNA levels even amid substantial female DNA background.[5,6] This study evaluated its performance with halved reaction volumes, aiming to increase sample throughput per kit while ensuring accuracy.

2. Material and methods:

2.1 Samples preparation

2.1.1 Control samples

Serial dilution of Control DNA (50 ng/µl) in the QuantiTect Nucleic Acid Dilution Buffer were prepared to construct a the standard curve including the following points:50ng/ul,1.8519ng/ul, 0.0686 ng/ul, 0.0025ng/ul + NTC.

The not-diluted control DNA and each dilution were also assayed as if it were known control sample.

2.1.2 Forensic sample

10 samples (oral swab, cigarette butt, bloodstain onto white cotton, bloodstain onblue denim, semen on toilet paper, vaginal swab, knife handle, hair root, 5-year femur, 5-year molar) were selected as representative of forensic samples. The femur and molar were from a male body exhumed five years after burial in a niche wall. All samples were taken in accordance with SIMEF laboratory internal procedures and informed consent.Samples were extracted, by the EZ1 Advanced XL DNA Investigator kit (Qiagen) using the biorobot EZ1 Advanced XL and the Trace protocol. [7].

DNA extraction from bones and teeth was performed according to a modified protocol previously published. [8]

2.2 Real Time Quantification

DNA Quantification was carried out using either the standard protocol recommended by the Manufacturer (9 µl of Quantiplex Pro RGQ Reaction Mix, 9 µl of Quantiplex Pro RGQ Primer Mix) than half reaction volumes, without altering the components ratio. 2 µl of control DNA dilutions were added to standard curve tubes, and 2 µl of unknown sample DNA were added to the individual sample tubes. [5]. Amplification was carried out in the Rotor-Gene Q cycler (Qiagen) according to the manufacturer cycling conditions. Control samples and forensic samples were run in triplicate using either regular then half reaction volumes.

2.3 Data analysis

Data obtained were analyzed by the software Q-Rex v.1 (Qiagen) using the Quant AssayData Handling Tool: quantification results, Inhibition Index (IC), Degradation Index (DI), Mixture Index (MI)were evaluated.[9] Accuracy was evaluated by determining the mean percentage error relative to the expected concentrations of the control samples. For precision assessment, the linear regression coefficient was calculated. All calculations, were carried out using a Microsoft Excel spreadsheet.

3. Results and Discussion

3.1 Standard curve

The standard curve is used for establishing the concentration of unknown samples in comparison with the standard ones.The R^2 value measures the fit of data to the regression line, with ≥0.990 being typical for standard curves. PCR efficiency,

indicated by the slope, ranges from –3.0 to –3.6, while the y-intercept reflects the Cq value for a 1 ng/µl sample. [5] Using both regular and half volumes, markers were efficiently amplified with R^2 (0.9986–0.9997), slopes (–3.03 to –3.16), and intercepts (21.06–21.58) All samples showed IC amplification, confirming no external inhibitors, and no amplification was observed for NTC in any channel.

3.2 Control samples

Each dilution of the control DNA used to construct the curve was assayed in triplicate as if it were control sample. Quantification accuracy assessed by calculating the difference between the observed and expected quantification values across four DNA input levels of the DNA control samples showed error values < 10%. Furthermore, we observed a good linear correlation between theoretical and measured concentrations for both human and male targets (R^2 = 0.9999 and 0.9997, respectively).

The mean percentage error related to the expected concentrations of the control samples using either regular than half volumes was < 10%, with the highest values respectively of 8,8% and 8%) at the lowest DNA concentration (0,0025ng/ul).

3.3 Forensic samples

Forensic samples, such as oral swabs, cigarette butts, bloodstains on white cotton or blue denim, semen on toilet paper, vaginal swabs, knife handles, hair roots, and femur or molar remains exhumed after five years, may be subjected to a wide range of environmental factors (e.g., heat, humidity, UV light) that can significantly affect DNA preservation and, consequently, the success of DNA profiling. These conditions can lead to DNA degradation, making the recovery of intact genetic material more challenging. [1] Moreover, the nature of the substrate on which the biological material is deposited also plays a crucial role in DNA extraction efficiency, with porous or textured surfaces often absorbing biological material, complicating retrieval efforts. In addition to environmental stressors, forensic samples may contain PCR inhibitors (i.e. heme, indigo dyes, soil contaminants) hat can interfere with amplification and reduce the quality or yield of DNA profiles. Severe inhibition can result in the loss of alleles from larger STR loci or produce completely false-negative results, a pattern often mistakenly attributed to significant template degradation. Mild to moderate inhibition may lead to partial allele loss and inaccurate estimation of the DNA quantity in the affected sample. [10] Overcoming these challenges requires optimized protocols and robust DNA quantification techniques to ensure accurate and reliable forensic analysis.

Furthermore, mixtures of DNA are common in forensic samples, particularly in cases of sexual assault. Detecting a mixed male-female DNA sample can be especially difficult when working with low-template DNA. In such cases, a mixed male-female sample may resemble a single-source male sample due to imbalanced amplification of autosomal and male-specific markers. The presence of a DNA mixture at the quantification stage is essential, as it enables the selection of the most appropriate analytical methods for accurate analysis. [11-13]

In the present study, differences in DNA quantity and quality were found among the analyzed forensic samples, due to the nature of substrates and the consequent different ability to preserve and to release biological material. (Table1)

Table 1: Quality assessment at Regular (R) and Half (H) Volumes

Samples		Mixture Index	Mixture Thr.	Degrad. Index	Degrad. Thr.	Male Degrad. Index	Male Degrad. Thr.	Inhib. Index	Inhib. Thr.
Oral swab	R	1,10	Bel.Thr.	1,16	Bel.Thr.	1,20	Bel.Thr.	0,17	Bel.Thr.
	H	1,26	Bel.Thr.	1,08	Bel.Thr.	1,22	Bel.Thr.	0,20	Bel.Thr.
Cigarette butt	R	1,23	Bel.Thr.	1,29	Bel.Thr.	1,15	Bel.Thr.	0,35	Bel.Thr.
	H	1,33	Bel.Thr.	1,16	Bel.Thr.	1,84	Bel.Thr.	0,27	Bel.Thr.
Bloodstain white cotton	R	1,11	Bel.Thr.	1,02	Bel.Thr.	1,10	Bel.Thr.	0,10	Bel.Thr.
	H	1,06	Bel.Thr.	1,05	Bel.Thr.	1,05	Bel.Thr.	0,14	Bel.Thr.
Bloodstain blue denim	R	1,12	Bel.Thr.	1,04	Bel.Thr.	1,24	Bel.Thr.	0,24	Bel.Thr.
	H	1,13	Bel.Thr.	1,07	Bel.Thr.	1,04	Bel.Thr.	0,18	Bel.Thr.
Semen toilet paper	R	1,05	Bel.Thr.	1,13	Bel.Thr.	1,12	Bel.Thr.	0,07	Bel.Thr.
	H	1,13	Bel.Thr.	1,16	Bel.Thr.	1,17	Bel.Thr.	0,10	Bel.Thr.
Vaginal swab	R	3,61	Mix	1,08	Bel.Thr.	0,31	Bel.Thr.	0,28	Bel.Thr.
	H	4,02	Mix	1,21	Bel.Thr.	1,06	Bel.Thr.	0,25	Bel.Thr.
Knife handle	R	2,58	Mix	1,21	Bel.Thr.	1,49	Bel.Thr.	0,42	Bel.Thr.
	H	2,71	Mix	1,13	Bel.Thr.	1,30	Bel.Thr.	0,36	Bel.Thr.
Hair root	R	1,14	Bel.Thr.	12,87	Degrad.	12,68	Degrad.	0,22	Bel.Thr.
	H	1,15	Bel.Thr.	13,44	Degrad.	13,13	Degrad.	0,17	Bel.Thr.
5 year Femur	R	1,22	Bel.Thr.	13,98	Degrad.	14,00	Degrad.	0,06	Bel.Thr.
	H	1,29	Bel.Thr.	14,73	Degrad.	14,27	Degrad.	0,10	Bel.Thr.
5 year Molar	R	1,08	Bel.Thr.	1,32	Bel.Thr.	1,20	Bel.Thr.	1,57	Inhib.
	H	1,23	Bel.Thr.	1,27	Bel.Thr.	1,28	Bel.Thr.	1,36	Inhib.

It was observed concordance per each sample between average quantification data- (from triplicate assays) obtained using regular and half reaction volumes. Generally, quantification values using half reaction volumes are almost higher than those obtained with regular volumes, with SD values ≤0,4. (Fig.1)

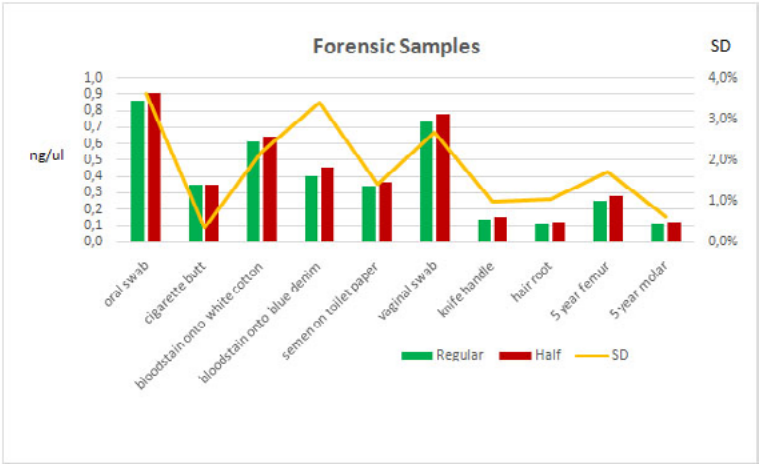


Fig.1. DNA concentration and standard deviation (SD) estimation for each sample

Quantification data analysis, using the Q-Rex software along with the Quant Assay Data Handling Tool, revealed (Table 1), in both cases,:

- Mixture Index (MI)>2, indicating a Female/Male mixture for the vaginal swab and the knife.
- Degradation Index (Human/Degradation) >10 for the hair root and 5-year femur
- Male Degradation Index >10 for the hair root and 5-year femur
- Inhibition Index (IC Shift)>1 for the 5-year molar

The default Degradation Indices of 10 enable STR profiling with DNA fragmented to 300 bp, differentiating fragments above or below this size. A Mixture Index above the default threshold of 2 suggests a potential male/female DNA mixture.

4. Conclusions

This study confirmed that the Investigator Quantiplex Pro RGQ Kit provides accurate and reliable results even with half the recommended reaction volumes. This

adjustment maintains analytical performance while increasing sample throughput and reducing reagent costs, a significant benefit in high-volume testing. The optimized protocol is especially valuable in resource-limited settings and large-scale studies, where cost-efficiency and processing capacity are critical. Importantly, reduced reagent volumes do not compromise the quality or reliability of forensic or clinical outcomes, making this a practical improvement to standard quantification protocols.

5. Conflict of interest statement

Nothing to declare.

6. References

[1] J.M Butler, Advanced Topics in Forensic DNA Typing: Methodology. Academic Press, 2012 Elsevier, ISBN 978-0-12-374513-2

[2] A.Barbaro, Challenges in DNA Extraction from Forensic Samples. in Dash, H.R., Shrivastava, P., Lorente, J.A,.(Eds) Handbook of DNA Profiling. Springer (Singapore), 2021; 1–20

[3] C.Schrader, A Schielke, L Ellerbroek et al. PCR inhibitors – occurrence, properties and removal. Journal of Applied Microbiology, 2012; 113:1014-1026

[4] C.Martins, G. Lima, M.R. Carvalho et al. DNA quantification by real-time PCR in different forensic samples, Forensic Science International: Genetics Supplement Series, 2015; 5:e545-e546

[5] Investigator Quantiplex Pro RGQ Kit Handbook QIAGEN Rev.1. 2018

[6] Evaluations of the Investigator Quantiplex ProRGQ Kit for DNA quantification and quality assessment of forensic casework sample, Application Note, QIAGEN,07/2018

[7] EZ1 DNA Investigator Handbook QIAGEN Rev. 7,2014.

[8] A.Barbaro, S. Samar, G. Falcone et al. Highly efficient and automated extraction of DNA from human remains using a modified EZ1 protocol. Forensic Sciences Research 2021;6(1):59–66

[9] Q-Rex Software User Manual, 01/2018

[10] R.Alaeddini. Forensic implications of PCR inhibition. A review, Forensic Science International: Genetics, 2012; 6 (3):297-305

[11] P.Gill P, C.H.Brenner, J.S. Buckleton et al. DNA commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures. Forensic Sci Int. 2006; 160:90–101.

[12] D.Taylor, J.A. Bright, J. Buckleton. The interpretation of single source and mixed DNA profiles. Forensic Sci Int Genet. 2013;7:516–28.

[13] N. Hu, B. Cong, S. Li, et al. Current developments in forensic interpretation of mixed DNA samples (Review). Biomed Rep. 2014; 2(3):309-316.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 865-870
DOI: https://dx.doi.org/10.15304/cc.2025.1869

STR Isoalleles Detection in Capillary Electrophoresis

Eugenia Carnevali[1*], Francesca Scarnicci[2], Ilaria Boschi[3], Martina Onofri[1,4], Simona Severini[1], Federica Tommolini[1], Maria Alessandra Marrucci[2], Vincenzo Lorenzo Pascali[2,3]

[1] *Forensic Sciences Laboratory, Section of Legal Medicine, Department of Medicine and Surgery, Santa Maria Hospital, University of Perugia, 05100 Terni, Italy*

[2] *Section of Legal Medicine, Department of Health Surveillance and Bioethics, Università Cattolica del Sacro Cuore, 00168, Rome, Italy*

[3] *Forensic Medicine operating Unit, Fondazione Policlinico Universitario IRCCS "A. Gemelli", 00168, Rome, Italy*

[4] *Section of Legal Medicine, Department of Medicine and Surgery, University of Perugia, 06123 Perugia, Italy*

** Corresponding author at: Azienda Ospedaliera "S.Maria" – Forensic Sciences Laboratory, via C. Mazzieri, 3, Terni, Italy,* e.carnevali@aospterni.it

Abstract

Two independent laboratories typed the X-STR profile of four samples that presented Off-Ladder alleles located outside the allele 23 bin at the DXS10135. The samples were amplified twice with the Investigator Argus X-12 QS-kit and for each replicate, four electrophoretic runs were performed with the SeqStudio™ Genetic Analyzer.

Two samples displayed a shift of around 0.42-0.48 bp from the reference sample, while two others displayed a shift of 0.30-0.32 bp. The four samples were sequenced using the ForenSeq DNA Signature Prep Kit on the MiSeq FGx Sequencing System. It emerged that the two samples with greater mobility shift shared the same isoallelic variant for the allele 23, characterised by an A→G transition in four repeats, leading to part of the sequence being [GAAA]15[GGAA]4 GAAA instead of [GAAA]20 (*wild-type* sequence). The other two samples presented the same A→G transition in three repeats (sequence [GAAA]16[GGAA]3 GAAA).

We suggest that the mobility shift for the observed allele 23 is due to the greater Molecular Weight of G compared to that of A, producing a fragment slightly heavier than the allele 23 of the allelic ladder. Our study demonstrates how, albeit not being as informative as Next Generation Sequencing techniques, more modern and sensitive Capillary Electrophoresis instruments, coupled with more stringent polymers, can assist in the detection of STR isoalleles.

Keywords
Off-Ladder Alleles; Isoallelic Variant; Capillary Electrophoresis; NGS

1. Introduction

Due to their higher discriminatory ability, STR isoalleles typing through Next Generation Sequencing (NGS) is becoming increasingly relevant in the forensic genetics field. However, in some instances, Capillary Electrophoresis (CE) may still provide valuable information [1-3].

For an inter-laboratory project, two independent laboratories typed the X-STR profile of the same sample (Stain.1) in replicates using CE technology. Both laboratories detected an Off-Ladder (OL) peak at locus DXS10135, characterised by a slightly higher number of bp than the respective allelic ladder peak for allele 23. The mobility shift observed was of around 0.34 – 0.47 bp (LAB 1) and 0.22 – 0.31 bp (LAB 2).

Considering that two independent entities, whose CE instruments have been internally validated and calibrated, encountered the same issue, that the shift was localised and was consistently observed across replicates, instrument-dependent problems were excluded. For confirmation, NGS was performed and it emerged that the OL allele observed at the DXS10135 locus of Stain.1 presented an A→G transition in four repeats compared to the reference, leading to part of the sequence being [GAAA]15[GGAA]4 GAAA instead of [GAAA]20 (*wild type* sequence).

In light of these observations, the present study aimed to analyse the electrophoretic mobility of samples with *wild type* (*wt*) and mutated alleles 23 at the DXS10135 locus. Therefore, the two laboratories sequenced with NGS technology and performed replicate PCR amplifications and CE runs of the X-STR profile of samples presenting peaks in or adjacent to the allele 23 bin at the DXS10135 locus.

2. Materials and Methods

The two laboratories re-analysed a pool of archive samples that presented the allele 23 or adjacent OL peak at the DXS10135 locus. In both laboratories, the same protocols and instruments used for analysing Stain.1 were employed. X-STR typing was performed using the Investigator Argus X-12 QS-kit (Promega) and fragments were separated using the SeqStudio™ Genetic Analyzer (Thermo Fisher Scientific).

Three samples with OL callings for allele 23 in DXS10135 were identified. The samples are herein referred to as Sample.1 to Sample.3.

In order to detect OL allele calling and to gauge electrophoretic mobility consistency across replicates, both labs PCR amplified each sample twice and each PCR replicate was injected four times, for a total of 16 data points. NGS typing was performed with the ForenSeq DNA Signature Prep Kit (Verogen) on the MiSeq FGx Sequencing System (Verogen).

3. Results and Discussion

Replicated STR typing and CE runs results are reported in **Table 1**. NGS typing results are reported in **Table 2**, in which, for reference, the results of a *wt* sample have been included. CE run examples are reported in **Figure 1**.

Table 1. X-STR PCR typing replicates. The allele or OL call is shown along with the number of bp measured for the peak

	LAB 1							
Sample	PCR replicate I				PCR replicate II			
wt Sample	23 - 309.40	23 - 309.39	23 - 309.40	23 - 309.40	23 - 309.27	23 - 309.36	23 - 309.32	23 - 309.37
Stain.1	OL - 309.64	OL - 309.69	OL - 309.69	OL - 309.77	OL - 309.77	OL - 309.77	OL - 309.69	OL - 309.73
Sample.1	OL - 309.77	OL - 309.77	OL - 309.76	OL - 309.86	OL - 309.77	OL - 309.81	OL - 309.77	OL - 309.77
Sample.2	23 - 309.42	OL - 309.76	23 - 309.69	23 - 309.69	23 - 309.60	23 - 309.69	23 - 309.60	23 - 309.65
Sample.3	OL - 309.82	OL - 309.76	23 - 309.69	23 - 309.69	23 - 309.69	23 - 309.65	23 - 309.69	23 - 309.65

	LAB 2							
Sample	PCR replicate I				PCR replicate II			
wt Sample	23 - 309.12	23 - 309.10	23 - 309.20	23 - 309.12	23 - 309.10	23 - 309.19	23 - 309.15	23 - 309.12
Stain.1	OL - 309.58	OL - 309.58	OL - 309.58	OL - 309.53	OL - 309.58	OL - 309.62	OL - 309.53	OL - 309.53
Sample.1	23 - 309.44	23 - 309.54	OL - 309.68	OL - 309.60	OL - 309.58	OL - 309.68	OL - 309.58	OL - 309.59
Sample.2	23 - 309.39	23 - 309.43	OL - 309.53	OL - 309.54	23 - 309.39	OL - 309.53	OL - 309.53	23 - 309.44
Sample.3	23 - 309.44	OL - 309.53	OL - 309.55	23 - 309.44	23 - 309.49	23 - 309.44	OL - 309.53	23 - 309.44

Table 2. The stretch of the sequence affected by the mutation is partially reported. «•••» represents the upstream and downstream portions of the STR not affected by the mutation.

Sample	Partial DNA Sequence
wt Sample	••• [GAAA]20 •••
Stain.1	••• [GAAA]15[GGAA]4 GAAA•••
Sample.1	••• [GAAA]15[GGAA]4 GAAA•••
Sample.2	••• [GAAA]16[GGAA]3 GAAA•••
Sample.3	••• [GAAA]16[GGAA]3 GAAA•••

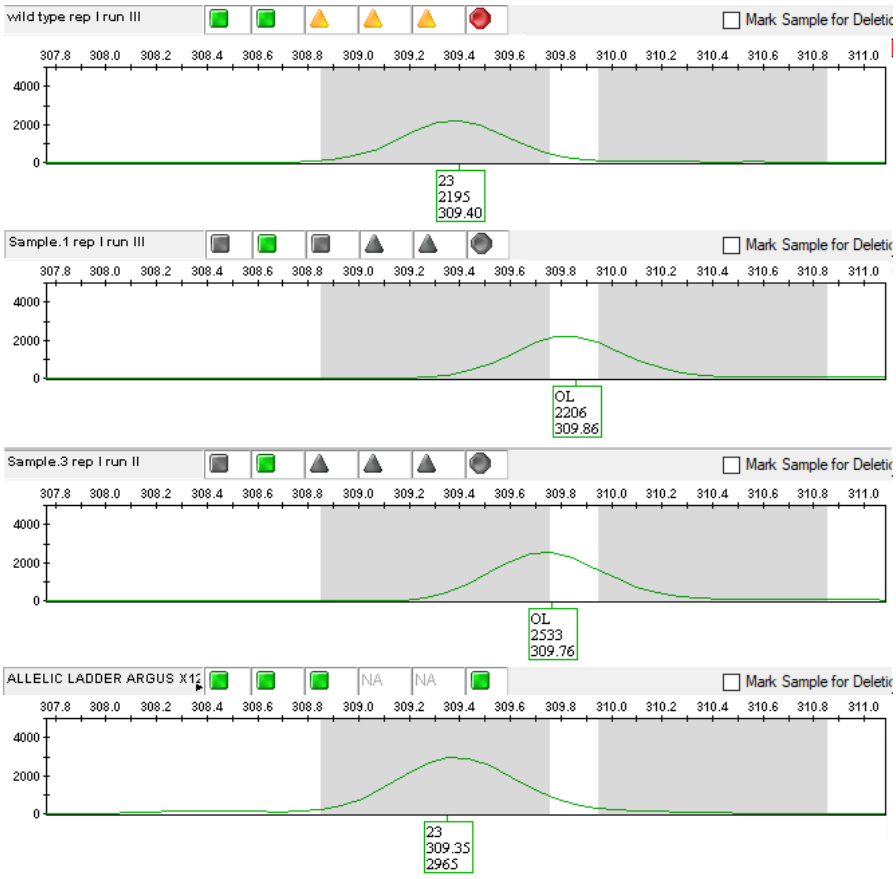


Figure 1. Electrophoretic mobility of the *wt* and mutated allele 23 at DXS10135. In descending other: wt Sample, Sample.1 (Mutation A), Sample.3 (Mutation B), Investigator Argus QX-12 Allelic Ladder of the run session.

As per NGS typing results, Sample.1 shared the same mutation as Stain.1 (mutation A), with partial sequence [GAA]15[GGAA]4GAAA, while Sample.2 and Sample.3 shared a second mutation (mutation B), with partial sequence [GAAA]16[GGAA]3GAAA.

Despite following the same protocol, a discrepancy of around 0.20 bp was observed between the results produced by the SeqStudio™ of the two laboratories. However, these slight variations between independent laboratories can be considered negligible if they are consistently observed. The calibration and internal validation of the instruments are a guarantee of reproducibility and replicability.

Overall, a difference in terms of molecular weight (expressed as bp) was observed between the mutated samples and the *wt* and between the two different mutations as well. Specifically:

— The mobility shift between mutation A samples (Stain.1 and Sample.1) and *wild type* was around 0.48 bp for LAB 1 and 0.42 bp for LAB 2.
— The mobility shift between mutation B samples (Sample.2 and Sample.3) and *wild type* was around 0.30 bp for LAB 1 and 0.32 bp for LAB 2.

In terms of detection:

— Mutation A allele 23 was flagged as an OL peak in 16/16 replicates for LAB1 and 14/16 replicates for LAB2 (87.5 - 100% of the cases).
— Mutation B allele 23 was flagged as an OL peak in 3/16 replicates for LAB1 and 7/16 replicates for LAB2 (18.75 – 43.75% of the cases).

Conclusions

We suggest that the mobility shift and greater number of bp observed for the mutated 23 alleles at the DXS10135 locus are attributable to the nature of the mutation (A→G transition) which produces slightly heavier fragments. In fact, the mobility shift is more pronounced and more frequently detected for mutation A samples, which have four mutated repeats, compared to mutation B samples, which only have three.

Our study demonstrates how, albeit not being as informative as NGS techniques, more modern and sensitive CE instruments, coupled with more stringent polymers, can assist in the detection of some STR isoalleles.

In fact, if the SeqStudio™ Genetic Analyzer is properly calibrated and internally validated and if the analysis is performed in replicate CE runs, some isoalleles

can be detected in up to 19 - 44% of the replicates, while others in up to 88 - 100% of the cases, depending on the extent of the mutation.

Funding

The authors did not receive external funding.

Declaration of competing interest

None.

References

[1] Gettings, K.B.; Aponte, R.A.; Vallone, P.M.; Butler, J.M. STR Allele Sequence Variation: Current Knowledge and Future Issues. *Forensic Science International: Genetics* 18 (2015) 118–130, doi.org/10.1016/j.fsigen.2015.06.005.

[2] Lee, E.Y.; Lee, H.Y.; Shin, K.-J. Off-Ladder Alleles Due to a Single Nucleotide Polymorphism in the Flanking Region at DYS481 Detected by the PowerPlex® Y23 System. *Forensic Science International: Genetics* 24 (2016) e7–e8, doi.org/10.1016/j.fsigen.2016.06.019

[3] Fujii, K.; Watahiki, H.; Mita, Y.; Iwashima, Y.; Miyaguchi, H.; Kitayama, T.; Nakahara, H.; Mizuno, N.; Sekiguchi, K. Next-Generation Sequencing Analysis of off-Ladder Alleles Due to Migration Shift Caused by Sequence Variation at D12S391 Locus. *Legal Medicine* 22 (2016) 62–67, doi.org/10.1016/j.legalmed.2016.08.003

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 871-878
DOI: https://dx.doi.org/10.15304/cc.2025.1869

MVCall: An automatic nanopore sequencing analysis pipeline for variation and heterogeneity identification of mitogenome

Jing Liu[1], Lirong Jiang[1], Hong Zhu[1], Linchuan Liao[2], Zheng Wang[1], Yiping Hou[1,*]

[1] *Institute of Forensic Medicine, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu 610041, China*

[2] *Department of Forensic Toxicological Analysis, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu 610041, China.*

[*] *Corresponding author at: Institute of Forensic Medicine, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu, China; email: forensic@scu.edu.cn (Y. Hou).*

Abstract

Mitochondrial DNA (mtDNA) analysis has gradually evolved from control region sequencing to whole genome sequencing with the development of massively parallel sequencing (MPS) technology, greatly promoting the application of mtDNA analysis in the fields of biomedicine, anthropology, and forensic science. However, the accuracy of MPS-based mtDNA analysis can be reduced due to uneven amplification and splicing of short fragments. Moreover, the heterogeneity resolution of MPS platforms is limited, approximately 10%. Nanopore sequencing technology has outstanding ability to sequence long continuous DNA fragments, making the detection of mitogenome without splicing possible. In this study, we developed an automatic analysis pipeline, named MVCall, based on nanopore sequencing data for identifying variation and heterogeneity in mitogenome. Compared with CmVCall, the mtDNA control region analysis pipeline built in our previous study, MVCall added an insertion/deletion (InDel) identification module, a consistency sequence generation module and a vcf file generation module, and optimized the analysis parameters of the polymer region. The sequence information and heterogeneity level of whole mitogenome can be obtained simultaneously by this pipeline. Overall, the MVCall provides a useful tool for mtDNA profiling in large-scale samples.

Keywords

Forensic genetics; Mitochondrial genome; MtDNA heterogeneity; Nanopore sequencing; MVCal

1. Introduction

The human mitochondrial DNA (mtDNA) is characterized by maternal inheritance, no recombination, and a higher mutation rate (in comparison with the nuclear genome), thus plays an important role in forensic genetics, population genetics, and anthropology [1-3]. Recent studies have shown that mtDNA heterogeneity, which refers to the coexistence of wild-type and mutant mtDNA in different tissues or cells within the same individual, is widespread in the population [4-6]. In the field of forensic genetics, mtDNA heterogeneity could be considered as another dimension of polymorphism and plays an essential role in certain cases, such as the identification of Romanov family's remains [7].

With the introduction and development of massively parallel sequencing (MPS) technology, mtDNA analysis has gradually evolved from control region sequencing to whole genome sequencing, and the amount of information obtained at one time has also increased [8]. However, MPS requires read splicing of multiple short fragment amplicons to attain the full mtDNA sequence, which increases the risk of heterogeneity assessment due to uneven amplification efficiency among amplicons and differential interference of nuclear mitochondrial sequences (NUMTs) [9-11]. Nanopore sequencing, as one of the mainstream third-generation long-read sequencing technologies, has outstanding ability for direct and real-time sequencing of nucleotide fragments of any length, making it possible to detect mitogenome without splicing [12].

In a previous study [13], an automated nanopore analytical pipeline (CmVCall) for the mtDNA control region was developed, and more point heterogeneity sites (PHPs) were displayed with a detection threshold of 5%. In the present study, we presented a long-amplicon sequencing approach targeting the human mitogenome within one amplicon and developed an automatic analysis pipeline for identifying variation and heterogeneity in mitogenome based on CmVCall, named MVCall. In addition, 92 peripheral blood samples were sequenced and analyzed through the above workflow to decipher the mtDNA polymorphisms, heterogeneity sites, and complex sites, thereby evaluating the applicability of this tool for mtDNA profiling in large-scale samples.

2. Material studied, methods, techniques

2.1 Sample preparation

This study was approved by the Ethics Committee of Sichuan University (Approval Number: K2019017) and performed according to the Declaration of Helsinki [14]. 92 peripheral blood samples were collected from healthy Han volunteers with written informed consent acquired from all participants. The QIAamp DNA Blood Mini Kit (QIAGEN) was utilized to isolate genomic DNA (gDNA) according to the manufacturer's recommendations. The NanoDrop 2000C Spectrophotometer (Thermo Fisher Scientific) was used to quantify the DNA quantity. Control DNA 9947A and 9948 (AGCU ScienTech Incorporation) were used as the positive control.

2.2. mtDNA whole genome amplification and QNome nanopore sequencing

The whole mitogenome of Control DNA (9947A and 9948) and 92 peripheral blood samples were amplified using the primer pair mt16426F/mt16425R and the Platinum SuperFi II PCR Master Mix (Thermo Fisher Scientific) according to the manufacturer's instructions. All PCR products were verified by 0.4% agarose gel electrophoresis and quantified utilizing the Qubit Fluorometer and the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). Nanopore sequencing of mtDNA was performed using QDR-V1.1 sequencing kit and Qcell-3841 chip (Qitan Technology) on a QNome-3841 sequencer, strictly following the manufacturer's protocol as described in the previous study [13].

2.3. MVCall development

Based on the previous study [13] and the nanopore sequencing data generated from the *Control* DNA, we developed a bioinformatic analytic pipeline, MVCall, for automated sequence designation and *heterogeneity* authentication of the whole human mitogenome. The overall pipeline is illustrated in **Fig.1**. Compared with the CmVCall [13], MVCall added modules for insertion and deletion identification, consensus sequence generation, and vcf file generation, and optimized the analysis parameters of the polymer region. Moreover, the effect of bases before and after the variant site was also assessed during the variant identification process.

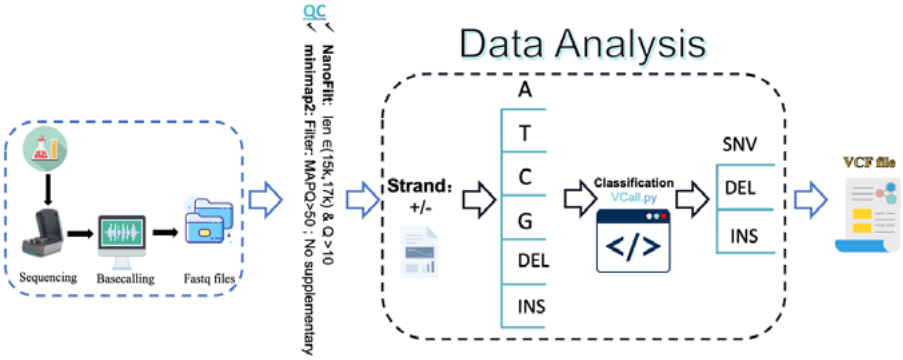


Fig. 1 MVCall pipeline. NiuTouGeng was used for base calling and to generate fastq files. Reads filtering was performed utilizing minimap2 and Nanofilt with filtering criteria of MAPQ > 50 and read lengths between 15 kb and 17 kb, respectively. Then, VCall.py was employed to authenticate heterogeneity sites and output vcf files including SNV, insertions, and deletions information.

2.4 Data processing

The mutational variants were manually verified utilizing the Integrative Genomics Viewer (IGV). Haplotype/haplogroup frequencies were calculated by the direct counting method, while the calculation of the haplotype/haplogroup diversity (HD) was as follows:

$$HD = \left[N \times \left(1 - \sum_{i=1}^{k} p_i^2\right)\right]/(N-1),$$

where *pi* is the frequency of the haplotype or haplogroup *i*, *k* means the number of haplotypes or haplogroups, and *N* represents the total sample size. According to the previous study [13], the heterogeneity authentication threshold was 5%, and the identified heterogeneity sites were compared with mutation sites reported in MITOMAP [15].

3. Results

3.1 QNome nanopore sequencing performance

The raw data was first filtered by length (15 kb to 17 kb) and Q-score (greater than 10). The average total reads of the 92 random peripheral blood samples was 4,263 and the average total bases was approximately 150 Mb after filtering. In addition, three replicate sequencing of Control DNA 9947A showed consistent results

in sequence identification, demonstrating the reproducibility and repeatability of nanopore sequencing. Interestingly, the variant at nt263 was not identified in any of the three sequencing attempts. The IGV results displayed that nt263 behaved inconsistently in both the forward and reverse strands and was therefore unavailable for judgment. As for the Control DNA 9948, the nt263 variant was not recognized due to the similar IGV performance as the Control DNA 9947A. With analogous IGV results as that of the nt263 variant, the identification of insertions (309.1C, 309.2C, and 315.1C of Control DNA 9947A, and 2156.1A and 8156.1G of Control DNA 9948) was relatively difficult, depending on the accuracy of sequencing platforms and mapping algorithms.

3.2 MtDNA genetic diversity

For whole *mitogenome, a* total of 85 haplotypes were detected in 92 samples, of which 79 were observed only once, giving a haplotype diversity of 0.9981. Meanwhile, a total of 72 haplogroups were detected in 92 samples, of which 58 haplogroups were unique, and the haplogroup diversity was 0.9936. When only sequence information of the control region was considered, 81 haplotypes were detected, of which 71 were observed only once, with a haplotype diversity of 0.9971. A total of 71 haplogroups were detected, of which 56 were observed only once, resulting in a haplogroup diversity of 0.9921.

3.3 Heterogeneity identification

Heterogeneity was detected at 372 loci using a detection threshold of 5.0%, with a total of 38 heterogeneity loci in the control region, accounting for approximately 10.22%, and 89.78% heterogeneity in the coding region. Compared with the mutation sites in the MITOMAP database, 123 (33.06%) heterogeneity sites were revealed with minor base mutations not reported in MITOMAP, and five heterogeneity-containing base insertion deletions were found in the coding region sites.

4. Discussion

Here, we present a long-amplicon sequencing method for the full length of the human mitogenome and develop an automatic bioinformatics pipeline for variation and heterogeneity identification of resulting long-read sequencing data. Our workflow successfully and efficiently recovered high-quality long-read sequencing data of the entire mitogenome. When analyzing Control DNA 9947A and 9948 sam-

ples, the nt263G variant and insertions at nt309, nt315, nt2156, and nt8156 were not accurately identified. After checking the IGV results and the rCRS reference sequence, we found that these sites are located in the polymer region, which has always been a challenge in sequencing analysis and requires the development of specialized algorithmic software. When analyzing the 92 random peripheral blood samples, we found that most of the identified variants were consistent with previously reported variants and that each individual carried different levels of heterogeneity. Furthermore, we observed about 33.06% heterogeneity not reported in the MITOMAP database, revealing the potential of nanopore sequencing to detect a wider spectrum of mtDNA mutations. Follow-up studies need to incorporate larger and more diverse sample sets, such as family tree samples, population samples, and samples from different tissues and body fluids, in order to fully explore the application of nanopore sequencing in mtDNA research and portray a more comprehensive mutation spectrum.

5. Conclusion

In this study, based on the nanopore sequencing data of the Control DNA 9947A and 9948 samples, we constructed MVCall, an automated variation identification and heterogeneity annotation system for mtDNA whole-genome long-read sequencing data. This system was then applied for mtDNA genome-wide analysis of 92 random peripheral blood samples. 85 haplotypes and 72 haplogroups were detected, with a haplotype and haplogroup diversity of 0.9981 and 0.9936, respectively. Using a detection threshold of 5.0%, a total of 372 heterogeneity sites were identified, and after comparing these heterozygous minor bases with the mutations reported in the MITOMAP database, 123 unreported heterozygous minor bases were identified, accounting for 33.06% of the total heterogeneity. Further studies should incorporate larger and more diverse sample sets to more reliably depict the nanopore sequencing-based mtDNA mutation profiles.

6. Acknowledgments

This study was supported by Sichuan Science and Technology Program (2024NSFSC0530), and Major Program of National Natural Science Foundation of China (82293653).

7. Conflict of interest statement

The authors declare that they have no conflict of interest.

8. References

[1] C. Yan, X. Duanmu, L. Zeng, B. Liu, Z. Song, Mitochondrial DNA: Distribution, Mutations, and Elimination, Cells 8 (4) (2019) 379.

[2] D. Syndercombe Court, Mitochondrial DNA in forensic use, Emerg. Top. Life Sci. 5 (3) (2021) 415–426.

[3] W. Kutanan, J. Kampuansai, M. Srikummool, A. Brunelli, S. Ghirotto, L. Arias, et al., Contrasting paternal and maternal genetic histories of thai and lao populations, Mol. Biol. Evol. 36 (2019) 1490–1506.

4] J.B. Stewart, P.F. Chinnery, Extreme heterogeneity of human mitochondrial DNA from organelles to populations, Nat. Rev. Genet. 22 (2) (2021) 106–118.

[5] B.A. Payne, I.J. Wilson, P. Yu-Wai-Man, J. Coxhead, D. Deehan, R. Horvath, et al., Universal heteroplasmy of human mitochondrial DNA, Hum. Mol. Genet. 22 (2) (2013) 384–390.

[6] K.M. Laricchia, N.J. Lake, N.A. Watts, M. Shand, A. Haessly, L. Gauthier, et al., Mitochondrial DNA variation across 56,434 individuals in gnomAD, Genome Res. 32 (3) (2022) 569–2582.

[7] P. Gill, P.L. Ivanov, C. Kimpton, R. Piercy, N. Benson, G. Tully, et al., Identification of the remains of the Romanov family by DNA analysis, Nat. Genet. 6 (2) (1994) 130–135.

[8] C. Børsting, N. Morling, Next generation sequencing and its applications in forensic genetics, Forensic Sci. Int. Genet. 18 (2015) 78–89.

[9] L. Albayrak, K. Khanipov, M. Pimenova, G. Golovko, M. Rojas, I. Pavlidis, et al., The ability of human nuclear DNA to cause false positive low-abundance heteroplasmy calls varies across the mitochondrial genome, BMC Genom. 17 (1) (2016) 1017.

[10] J.L. O'Donnell, R.P. Kelly, N.C. Lowell, J.A. Port, Indexed PCR Primers Induce Template-Specific Bias in Large-Scale DNA Sequencing Studies, PLoS One 11 (3) (2016) e0148698.

[11] W. Wei, K.R. Schon, G. Elgar, A. Orioli, M. Tanguy, A. Giess, et al., Nuclear-embedded mitochondrial DNA sequences in 66,083 human genomes, Nature 611 (7934) (2022) 105–114.

[12] J. Clarke, H.C. Wu, L. Jayasinghe, A. Patel, S. Reid, H. Bayley, Continuous base identification for single-molecule nanopore DNA sequencing, Nat. Nanotechnol. 4 (4) (2009) 265–270.

[13] L. Jiang, J. Liu, S. Li, Y. Wen, X. Zheng, L. Qin, et al., CmVCall: An automated and adjustable nanopore analysis pipeline for heteroplasmy detection of the control region in human mitochondrial genome, Forensic Sci. Int. Genet. 67 (2023), 102930.

[14] A. Nicogossian, O. Kloiber, B. Stabile, The Revised World Medical Association's Declaration of Helsinki 2013: enhancing the Protection of Human Research Subjects and Empowering Ethics Review Committees, World Med. Health Policy 6 (2014) 1–3.

[15] M.T. Lott, J.N. Leipzig, O. Derbeneva, H.M. Xie, D. Chalkia, M. Sarmady, et al., mtDNA Variation and Analysis Using Mitomap and Mitomaster, Curr. Protoc. Bioinformatics 44 (123) (2013) 1.23.1–1.23.26.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 879-885
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Advancing Semen Detection Techniques: Utilizing STK® Sperm Tracker to Emphasize Traces and Potentially Define Time Since Deposition (TSD)

Selena Cisana[1*], Monica Omedei[1], Paolo Garofano[1], Aurora Desogus[2], Giorgia Ferrero[2], Florian Tharin[3], Eugenio Alladio[1,2*]

[1]*Centro Regionale Antidoping "A. Bertinaria", Orbassano (Turin), Italy*
[2]*Department of Chemistry - University of Turin (Italy)*
[3]*AXO Science, Villeurbanne (France),*

** Corresponding author at: selena.cisana@gmail.com; eugenio.alladio@unito.it*

Abstract
Recently a new innovative technology has emerged for semen detection: STK®Sperm Tracker (AXO Science). This reagent is a non-toxic presumptive test specifically designed to highlight male acid phosphatase (PA) in human seminal fluid. Upon interaction with this enzyme, the reagent triggers a reaction that enables the observation of persistent blue fluorescence, after subsequent exposure to UV-365 nm lamp irradiation. This study aims to evaluate the efficiency of STK® in emphasizing and detecting semen traces across various scenarios and environments. Additionally, interesting time-dependent color changes were observed, suggesting a potential technique for Time Since Deposition (TSD) estimation.

Keywords
Semen, Principal Component Analysis (PCA), Time Since Deposition (TSD)

1. Introduction

In forensics, detecting body fluids at crime scenes or on evidence is critical for proper evidence collection [1]. Various sensitive, specific, and non-destructive tests have been developed for different types of fluids. Recently AXO Science (Villeurbanne, France) introduced a new technology for semen detection: Sperm Tracker® [2]. This non-toxic, ready-to-use technology enhances both the sensitivity and

specificity of semen detection, offering significant advantages in locating and collecting semen stains. This study assesses the effectiveness of STK® in highlighting and detecting semen traces across various scenarios and environments. The technique preserves downstream DNA integrity without any negative effects, thus, significantly improving trace detection. Exploratory analysis using pixel-level and image-level approaches on RGB raw data, combined with Principal Component Analysis (PCA), underscores the method's practical utility in identifying sperm traces.

2. Material studied, methods, techniques

2.1. Material I

To test the STK® Skin (a sprayable solution that allows to specifically detect semen on victim's skin and hair), mock samples were prepared using pig ears and both natural and bleached human hairs. Fresh sperm was collected from at least 3 donors to reduce interindividual variability, and a 50 µL was laid down in triplicate on each surface. The samples were then kept under different environmental conditions (wet, wet outdoor, room temperature and controlled temperature at 40°C) for different storage times (20 days, 10 days, 5 days, 3 days, 2 days, 24 hours, 12 hours and 1 hour). After STK®Skin treatment, the emphasized traces were photographed in a standardized manner using VILBER VL-6.L UV lamp (365nm) for exposure: Nikon digital camera D7100 with a DX AF-S NIKKOR 10-24mm 1:3.5-4.5G ED camera lens with yellow filter was positioned on a tripod, 15 cm far from samples with 1 second capture time.

Subsequently, exploratory analyses, including pixel-level and image-level approaches on RGB raw data, along with Principal Component Analysis (PCA) [3], were conducted by computing the colorgramns on the RGB collected images [4].

2.2. Material II

To test the STK® Spray (a presumptive test for the detection of human male semen, intend to be used indoor and outdoor, on floor, soil, hard surfaces like furniture, plastics, leather, metal, leaves, etc), double-blind experiments were carried out: both Non-Expert and Expert operators participated to compare accuracy and sensitivity of STK®Spray in detecting traces of semen on different surfaces. Fresh sperm samples were collected from at least 3 donors and varying volume of samples were spotted on different surface. The detection was carried out at two different times: immediately after deposition and seven days after deposition.

In addition, 50 µL of semen was spotted in quintuplicate onto rigid blue-plastic supports (to ensure consistent material and color reference) and stored at room temperature. The evidences were then emphasized with STK®Spray at 4 different times since deposition (1-hour, 1 month, 2 months and 3 months). After the STK® Spray treatment, the highlighted traces were photographed in a standardized manner, and image-based approach was adopted and subsequently analyzed using PCR to conduct exploratory analysis of the image dataset.

3. Results

The difference observed before and after the application of STK® Skin is remarkable. After the STK® application, the areas of interest become well-defined, significantly enhancing evidence collection.

To highlight the presence of semen traces, a PCA model was built for each captured image. An example is reported in Fig. 1: each graph shows PC1/PC2 and PC3 scores and each point represents a pixel from the original image. The color reflects pixel density, where red shows high density and blue shows low density. The PC1/PC2 scores plot effectively differentiates seminal fluid areas from the background. Background scores cluster on the left at low PC1 and PC2 values, while trace pixels form a distinct cluster on the right at high PC1 and PC2 values.

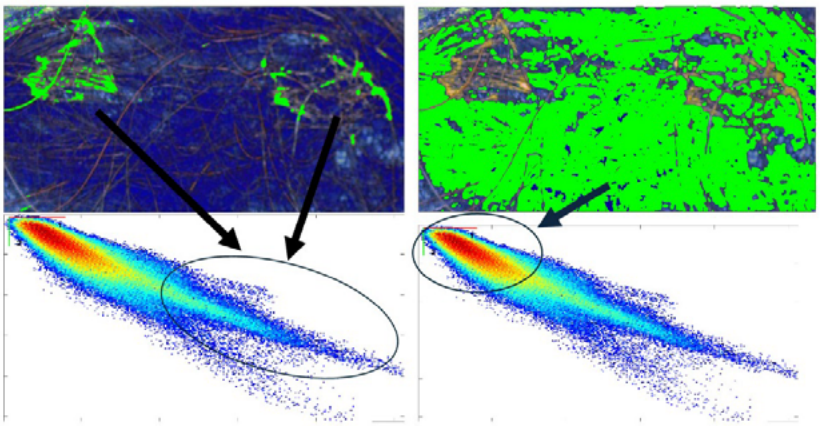


Figure 1. Semen traces on hair: selected image pixels shown as different clusters on PC1/PC2 scores.

To detect the actual effectiveness of the STK® Spray under the tested conditions, the graph in Fig. 2 shows the percentage number of traces detected, before and after treatment, relative to the total traces deposited, correlated to operator's experience (Fig. 2A). Furthermore, no significant variations in stain detection capability are revealed concerning trace volume (Fig. 2B) or Time Since Deposition (TSD) (Fig. 2C).

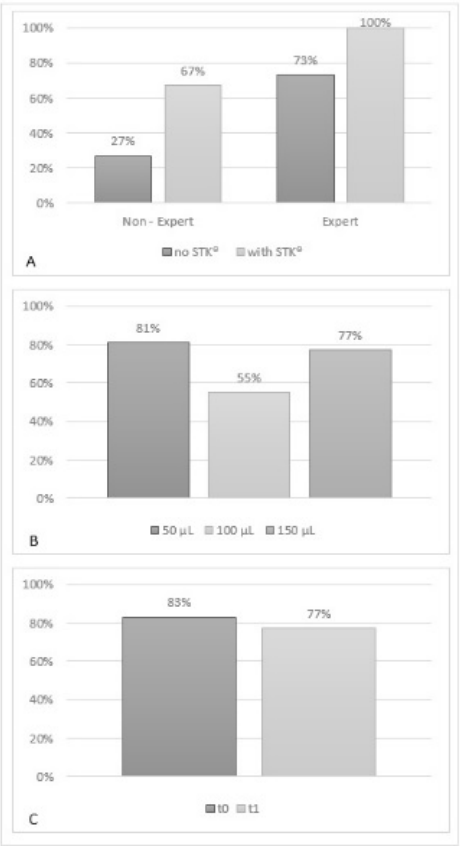


Figure 2. A) The graph compares the percentages of traces positively considering the detection solely with ALS (Alternative Light Source) and after STK® Spray treatment, in correlation between Non-Expert and Expert operators. **B)** The graph comparison of the percentage of different volumes traces (50 µL, 100 µL, 150 µL) correctly detected. **C)** Comparative graph of the percentage of semen both freshly (t0) and after 7 days of TSD (t1) detection.

During the chemometric analysis, an exciting colorimetric change in STK® fluorescence has been revealed over TSD, prompting further investigation. The matrix of signals derived from trace images at different deposition time detection was analyzed using PCA method, revealing TSD as a significant source of variability. The loadings of PC1 indicate that the intensity of the red channel plays a crucial role: low red values result in positive loadings, clustering samples at t0 (1-hour) in positive PC1. The trend was also observed in RGB values comparison [5], as shown in Fig. 3: chromatic variation is due to increased red index over time.

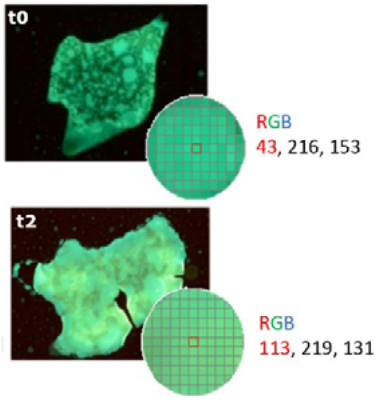


Figure 3. RGB Color Picker interface for specific pixel provides the RGB channels' intensities. 1-hour (t0) and 2 months (t 2) detection TSD are compared.

4. Discussion

The use of STK® technologies allows an effective increase in semen detection. The PCA analysis aims to differentiate trace-related areas from irrelevant background. However, the analysis showed different results between pig ears and hair. STK® Skin was less effective on hair, probably because the trace dispersed throughout the hairs, unlike skin samples. RGB imaging only considers color properties, making identifying traces treated with STK® Skin difficult if their color is too similar to the rest of the sample. More advanced imaging systems like hyperspectral imaging could be used to implement the analysis.

The stains' volume and the detection time since deposition do not significantly influence the detection of traces with STK®Spray. It would be interesting to explore further using minimal volumes of semen traces or mixing them with other biological fluids. Time-dependent trace color changes were also noted, suggesting

potential aging indicators: This variation over time could be determined by a decrease in the activity of acid phosphatase, the target of the reagent.

5. Conclusion

The STK® has proven to be more efficient in semen detection, than the examination conducted solely with Alternative Light Sources (ALS). A significant reduction in the risk of false positives has also been observed. The STK® significantly helps detection even for non-expert operators emphasizing its easy-use and effectiveness in semen detection.

Exploratory analysis used pixel and image-level approaches revealing consistent efficiency across varying environmental conditions. TSD has been revealed as a significant source of variability: this would to develop a multivariate regression model that can estimate the time since deposition of an unknown trace.

6. Acknowledgments

Special thanks to our donors, for the constant availability and the unconditional support.

7. Conflict of interest statement

This presentation is supported by AXO Science in the form of a non-profit collaboration for research purposes only. AXO Science provided no monetary funds.

8. References

[1] T. Sijen, S. Harbison, *On the Identification of Body Fluids and Tissues: A Crucial Link in the Investigation and Solution of Crime,* Genes (Basel) 12 (2021). doi.org/10.3390/genes12111728.

[2] E. Borges, A. Degiuli, S. Desrentes S., *et al*., *Evaluation of the SPERM TRACKER™ for Semen Stains Localization on Fabrics*, J. Forensic Res. 8 (2017). doi.org/10.4172/2157-7145.1000380.

[3] S. L. C. Ferreira, *Chemometrics and Statistics | Experimental Design*, Encyclopedia of Analytical Science (Third Edition), Academic Press, 2019. https://doi.org/10.1016/B978-0-12-409547-2.14536-6.

[4] D. Wu, D.-W. Sun, Colour measurements by computer vision for food quality control – A review, Trends Food Sci. Technol. 29 (2013). doi: https://doi.org/10.1016/j.tifs.2012.08.004.

[5] R. Calvini, G. Orlandi, G. Foca, *et al*., *Colourgrams GUI: A graphical user-friendly interface for the analysis of large datasets of RGB images*, Chemom. Intell. Lab. Syst. 196 (2020). doi.org/10.1016/j.chemolab.2019.103915.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 887-890
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Development of an automated workflow for library preparation of ForenSeq assays

Mario Scherer[1*], Miroslav Vraneš[1], Margaretha Willuweit[1], Stefan Cornelius[1], Daniele Terramagra[1], Christian Pustelny[1]

[1] *QIAGEN, Hilden, Germany.*

** Corresponding author at: QIAGEN, QIAGEN Strasse 1, 40724 Hilden, Germany, mario.scherer@qiagen.com, +49-15201811486.*

Abstract

Massively parallel sequencing (MPS) provides more information on samples, which helps to make identifications in challenging cases. The ForenSeq MainstAY kit combines autosomal and Y-STR loci into a comprehensive panel of 52 STRs and Amelogenin. The generation of MainstAY sequencing libraries is a multistep process involving magnetic bead based procedures for purification and normalization. Here we describe the development of an automated workflow, based on the EZ2 Connect Fx and the QIAgility. In this workflow, the QIAgility can be used to perform the PCR setup and library pooling steps. The EZ2 Connect Fx is used for library purification and normalization in a single protocol. The prefilled EZ2 cartridges increase reproducibility and significantly reduces hands-on time. Sensitivity and reproducibility of the automated workflow were compared to manual library generation. The total number of reads obtained, as well as allele recovery were found to be equivalent. The risk of cross-contamination was studied by running alternating positive and negative samples. Across 24 negative samples, two alleles above threshold were detected.

Keywords

MPS, automation, library preparation, ForenSeq, EZ2.

1. Introduction

Compared to historic methods like capillary electrophoresis (CE), massively parallel sequencing (MPS) enables the simultaneous analysis of more genetic markers, and provides a higher detailed genetic profile. The ForenSeq MainstAY kit

used in this study amplifies 27 autosomal STRs, 25 Y-STR loci and Amelogenin. However manual library preparation for sequencing is an obstacle and broader adoption for routine use, needs affordable and convenient automation for low to

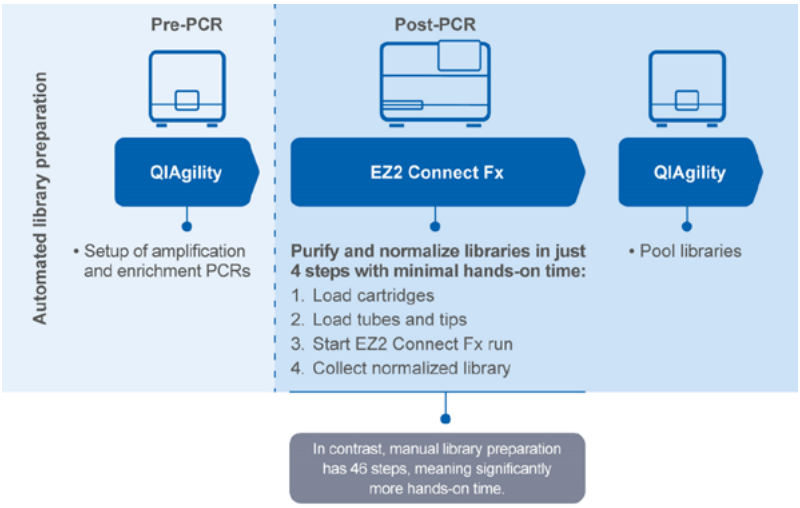


Figure 1. Workflow overview.

mid throughput. In particular the magnetic bead based purification and normalization steps of the workflow are tedious. The EZ2 Connect Fx instrument is designed for automated magnetic bead processing and can run 24 samples at a time. The instrument uses prefilled reagent cartridges. To complete the automated workflow, PCR reaction setup and library pooling steps can be performed by the QIAgility, or another suitable liquid-handling instrument.

2. Methods

For manually prepared ForenSeq MainstAY libraries, the reference guide of the kit was followed. The automated library purification and normalization was performed using preliminary research protocols on the EZ2 Connect Fx instrument. PCR amplification was done on Veriti or Proflex 96-well PCR systems (Thermo Fisher Scientific). Positive control DNA NA24385 was used as template and serially diluted for sensitivity studies. For studies evaluating the reproducibility of the automated library purification, enriched PCR products were pooled and aliquots

of the pool used for manual, or automated procedures. Sequencing was performed using the MiSeq FGx Reagent Micro Kit. Data was analyzed with the Universal Analysis Software 2.0 and the default ForenSeq MainstAY analysis method.

3. Results & discussion

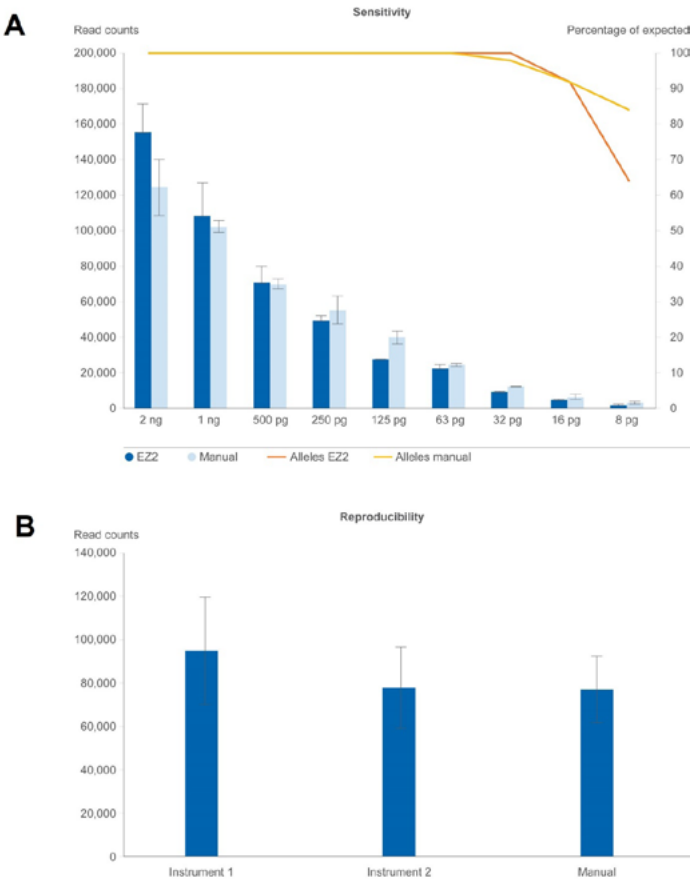


Figure 2. Sensitivity and equivalence to a manual library purification and normalization. A) Serial dilution of control DNA (NA24385) was amplified and library purification and normalization done manually, or with the EZ2 Connect Fx. B) 1 ng control DNA (NA24385) was amplified and library purification and normalization done manually, and with two EZ2 Connect Fx instruments. 24 replicates each were run.

In order to test for sensitivity and equivalence to a manual library purification and normalization, positive control DNA (NA24385) was serially diluted and amplified in 4 replicates each using the ForenSeq MainstAY Library Prep Kit. One half of the amplified libraries was purified and normalized manually according to the handbook protocol, while the other half was run on the EZ2 Connect Fx. All normalized libraries were pooled, sequenced on the MiSeq FGx, and analyzed using the Universal Analysis Software. Results were equivalent for the manual and automated workflows. Consistent full profiles were obtained down to 63 pg, lower input amounts showed allele dropouts as expected (Fig. 2A).

The reproducibility across different EZ2 Connect Fx instruments was addressed by running 15 replicates each on two instruments against a manual control. 5 µL of each library was pooled for sequencing. Average read counts and sample-to-sample variation were in the expected range, and comparable between manual and automated procedures (Fig. 2B). All samples resulted in full ForenSeq MainstAY profiles.

A cross-contamination study was performed. Positive control DNA (NA24385) was amplified using 1 ng template. Positive samples and NTCs were handled separately during amplification. However, spurious contamination observed may still have been introduced during PCR of negative controls. Amplified libraries were purified on the EZ2 Connect Fx instrument alternating between positive & negative samples. Two runs on different instruments were performed. All positive samples resulted in the expected full profiles with all 53 markers detected with the expected profiles. In 2 of 24 NTCs, a single allele was detected above threshold (data not shown). One allele did not match the control DNA profile, thus could be excluded to be a cross-contamination from the automated run. The second allele matched the control DNA profile. The rate of sporadic background observed was similar to that of manually prepared libraries. In general, EZ2 Connect Fx instruments used for library clean up should be operated in post PCR areas, and not process sample extraction.

4. Conclusion

Preliminary results suggest the EZ2 Connect Fx is suitable for ForenSeq library purification. The automated protocol provides equivalent sensitivity compared to a manual library preparation, free of cross contamination.

5. Conflict of interest statement

All authors are QIAGEN employees.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 891-898
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Forensic validation of the MPS Forenseq MainstAY assay

Amaia Cabrejas-Olalla[1], María de la Puente[1], Ana Freire-Aradas[1], Adrián Ambroa-Conde[1], Jorge Ruiz-Ramírez[1], Javier González-Bao[1], Lucía Casanova-Adán[1], Miguel Boullón-Cassau[1], Amelia Rodríguez-López[1], Christopher Phillips[1,2], María Victoria Lareu[1], Ana Mosquera-Miguel[1*]

[1]*Forensic Genetics Unit, Institute of Forensic Sciences, Universidade de Santiago de Compostela, Spain.*

[2]*King's Forensics, Faculty of Life Sciences and Medicine, King's College, London, UK.*

** Corresponding author at: ana.mosquera@usc.es; +34 881812404*

Abstract

Massively Parallel Sequencing (MPS) has emerged as a powerful tool for overcoming some of the current challenges of forensic genetics thanks to its improved sensitivity, higher discrimination power due to the capacity to evaluate marker's sequence variation and enhanced mixture interpretation. However, despite the increasing accessibility of MPS assays and platforms, results from forensic internal validations are not usually available in the literature, so it is difficult to evaluate the performance of STR analysis with MPS in real forensic routine casework.

The Forenseq® MainstAY SE assay is composed of 28 autosomal STRs (aSTRs), including the highly polymorphic SE33 marker, and 25 Y-chromosome STRs (Y-STRs), offering the benefits of combined analysis of autosomal and Y STRs in one assay, and expanding the capabilities of the current 6-dye capillary electrophoresis kits. Here we present a forensic validation pipeline for the MainstAY SE assay, and we evaluate its performance on forensic routine cases. Protocol optimization using reduced volumes has been explored. Sensitivity analyses using dilution series, common control sample capillary electrophoresis-MPS genotype concordance evaluations, and analysis of artificial mixtures with different proportions of DNA from several donors. These evaluations have allowed us to define the limits of the MainstAY assay and to identify underperforming markers. Sequence coverage, stutter ratios, heterozygous balance, length/sequence variation and stacking artifacts were among the parameters evaluated for determining individual STR performance.

Keywords

Massively Parallel Sequencing (MPS), internal validation, reduced volumes, sensitivity, kinship cases, DNA mixtures.

1. Introduction

Massively Parallel Sequencing (MPS) has emerged as a powerful tool to overcome some of the current challenges in Forensic Genetics due to its improved sensitivity, higher discrimination power from the capacity to evaluate component marker's sequence variation, and enhanced mixture interpretation [1]. However, despite the increasing accessibility of MPS assays, results from forensic internal validations are not easily available, making it difficult to evaluate the performance of STR analysis with MPS in real forensic routine casework.

The Forenseq® MainstAY SE assay [2] is composed of 28 autosomal STRs (aSTRs), including the highly polymorphic SE33 marker, and 25 Y-chromosome STRs (Y-STR), offering the benefits of combined analysis of autosomal and Y STRs in one assay, and expanding the capabilities of the current 6-dye capillary electrophoresis (CE) STR kits.

Here we present a forensic validation pipeline for the Forenseq® MainstAY SE assay (Verogen-Qiagen), to evaluate its performance for its application to forensic casework.

2. Material and methods

2.1. Samples included and assay optimization

The sample set consisted of NA24385 and 007 DNA controls, that were used for serial dilution replicates (1, 0.5, 0.25, 0.125, 0.0625 and 0.0312 ng/mL), reduced volume experiments and the preparation of artificial DNA mixtures at different ratios (1:1, 1:3, 1:9, 1:14 and 1:24; being NA24385 the major contributor and CP007 the minor one). Eight reference samples, obtained from voluntary donors with informed consent approved by the Bioethics Committee of the Universidade de Santiago de Compostela (USC-58/2022), were extracted with PrepFiler Express (Applied Biosystems, AB) and quantified Quantifiler™ Trio DNA Quantification Kit (AB). For CE-MPS concordance purposes, samples were analyzed with capillary electrophoresis STR kits, namely GlobalFiler (AB) [3], and YfilerPlus (AB)[4]

Library preparation was carried out following the manufacturer protocol (Forenseq® MainstAY Reference Guide - Verogen-Qiagen). For the assay optimization, replicates of 1ng and 0.5ng DNA inputs of 007 control DNA, reducing the reaction mix volumes (half and third), were run. Diluted pooled libraries were sequenced in a MiSeq FGx™ system following the equipment instructions. Results were analysed with the ForenSeq UAS software v.2.5.0.

2.2. Forensic validation

Validation studies were performed following the recommendations of the Scientific Working Group on DNA Methods (SWGDAM) [5], consisted of the evaluation of CE-MPS concordance, reproducibility and sensitivity studies (performed with the dilution series replicates), performance in forensic samples and mixture analysis.

Familias3 [6] software was used for Random Match Probability (RMP) statistical evaluation, and EuroForMix software [7] was employed for mixture analysis. Published allele frequencies from the population of interest were selected for the different STR panels (GlobalFiler™ [8, 9] and MainstAY SE [10]).

3. Results

3.1. Protocol optimization with reduced reaction volumes

As a first step and aiming to achieve the maximum cost-effectiveness of the MPS assay, the protocol optimization consisted of the evaluation of the possibility to reduce the reaction volume, by comparing the results from full, half and third volumes.

Mean concordance and profile completeness percentages for full and reduced volumes were calculated, and full profiles and 99-100% CE-MPS concordance were achieved for Y-STRs and aSTRs, respectively, when using the manufacturer recommended volume (Figure 1). However, considerable coverage variability was detected for some of the reduced volume replicates. Percentage of completeness for half volume replicates varied between 88-100% (mean of 97%) for aSTRs and 52-92% (mean of 81%) for Y-STRs. A significant decay was appreciated for the third volume samples, with 0-57% (mean:15%) for aSTRs and 0-32% (mean: 9%) for Y-STRs percentage of completeness. Y chromosome profiles seem to be more affected by the reduction volume strategy. Despite these values, the concordance percentage was considerable (>87%) even for very partial profiles.

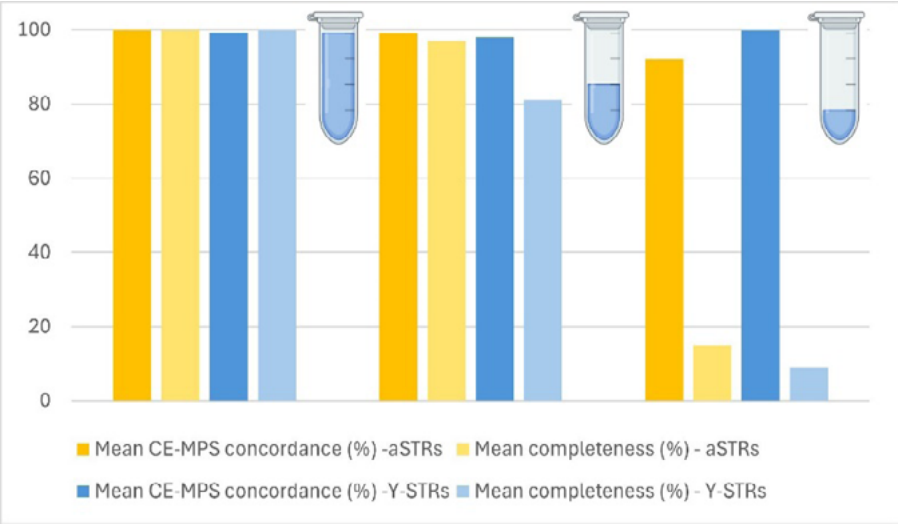


Figure 1. Mean CE-MPS concordance and mean profile completeness percentages for full, half and third volumes.

3.2. Sensitivity and concordance

Dilution series replicates were run to establish the limits of the assay. Satisfactory results in terms of reproducibility and mean profile completeness (>94%) were obtained up to 0.125 ng DNA input.

Generally, concordant results have been achieved for shared autosomal STRs and Y-STRs between CE and MPS-length data, with some profile instability increases (84% and 90% concordance for aSTRs and Y-STRs, respectively) in the lowest DNA inputs.

3.3. MPS panel technical performance

Considerable coverage variability has been noticed in sample replicates, not only among, but also within sequencing runs, especially with decreasing DNA input. Some underperforming markers were identified, namely Penta E, Penta D, D10S1248, D19S433, D2S1338 and D12S391 (Figure 2). Frequently, these markers were not genotyped in the most compromised samples.

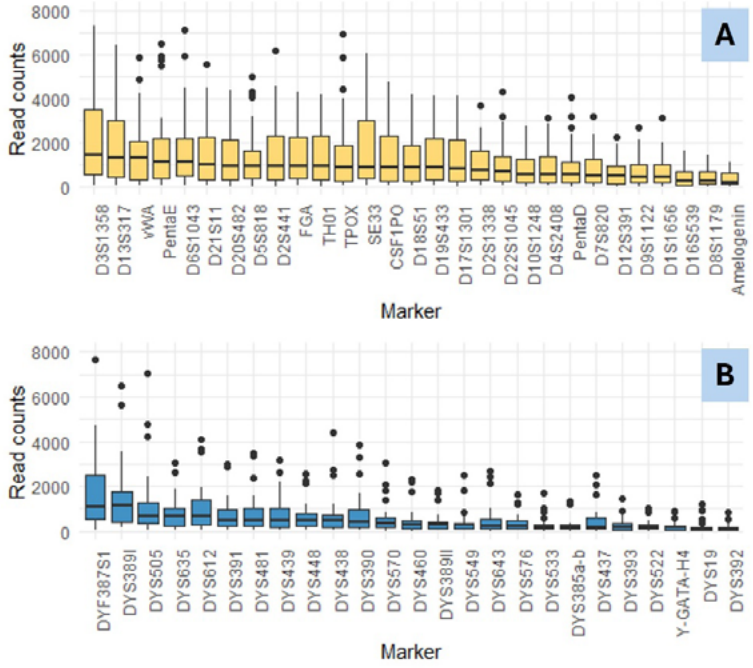


Figure 2. Coverage variation observed for aSTRs (A) and Y-STRs (B) among the ten single source samples analyzed.

Heterozygous imbalance was found even in good quality samples, being significantly pronounced for Penta E, SE33, D13S317, D22S1045, D16S539 and DYS385. However, it only compromised profile interpretation in low quality samples or mixtures.

High stutter ratios (>20%) are the most frequently observed artifacts in the panel, affecting both aSTRs (D1S1656, FGA, D2S1338, D8S1179, D12S391) and Y-STRs (DYS612 and DYS481). Sequencing dependent artifacts, such as stacking, could be solved by repeating the sequencing step.

3.4. Forensic informativeness

For the analyzed eight unrelated single source samples, 10 aSTRs presented isoalleles (e.g. D9S1122, D21S11, D13S317, D12S391, …). The number of isoalleles per sample ranged between 0 and 5, however MPS sequence data can have a great impact on the RMP. For 3 reference samples RMP were calculated using length or sequence MPS data, obtaining differences between 8-20 orders of magnitude.

Artificial mixtures were analyzed to determine the power of MainstAY panel for the identification of DNA contributors in mixed profiles. Even for the most imbalanced mixtures, 1:14 and 1:24, 15 and 9 minor contributor exclusive alleles were detected, respectively. For 1:1 and 1:9 artificial mixtures Likelihood Ratio (LR) was calculated using length or sequence MPS data. Notable increase of the logLR value is observed in all cases when using sequence information. For imbalanced mixtures (1:9), this effect was more pronounced for minor contributor.

4. Discussion and conclusions

The present work aimed to evaluate and validate the Forenseq® MainstAY SE assay to be applied to forensic casework. With this goal in mind, a forensic validation was performed. The assay performance was satisfactory for both reference and good quality samples, presenting full concordance and profile completeness in most of the samples and up to 0.125 pg. However, manual revision with a more stringent criterion should be applied (e.g. higher analytical threshold) when dealing with low quantity DNA or degraded samples and mixtures. Sequence information and detection of isoalleles can greatly improve the interpretation of real casework mixtures; however, additional analyses are required.

The evaluated panel, offering the benefits of combined analysis of autosomal and Y-STRs in one assay, can contribute to the resolution of non-direct kinship cases. However, to take advantage of the sequence MPS data, the UAS software and the presentation of the results should be improved to alleviate the complexity of the analysis. Moreover, additional efforts in scalability and technical improvements (e.g. stutter ratio) would allow forensic labs with low-medium casework loads to implement not only kinship analysis but also criminal casework.

5. Acknowledgments

MdP is supported by grant IJC2020–042638-I funded by the Gobierno de España MCIN/AEI/10.13039/501100011033 and the European Union "NextGenerationEU/PRTR". JRR is supported by the "Programa de axudas á etapa predoutoral" funded by the Consellería de Cultura, Educación e Ordenación Universitaria e da Consellería de Economía, Emprego e Industria from Xunta de Galicia, Spain (ED481A-2020/039). JGB is supported by a predoctoral grant (reference FPU22/02683) funded by

Ministerio de Universidades from Gobierno de España. LCA is supported by a predoctoral grant funded by "Fundación Caixa Rural Galega Tomás Notario Vacas". MBC is supported by the "Programa de axudas á etapa predoutoral" funded by the Consellería de Cultura, Educación e Formación Profesional e Universitaria from Xunta de Galicia (ED481A-2024-163). MVL is IP of the Project PID2019–107876RB-I00 funded by the Ministerio de Educación, Cultura y Ciencia, Spain (MCIN/AEI/ 10.13039/501100011033) and of the Project PID2022-141224OB-I00 funded by MCIN/AEI/ 10.13039/501100011033 and "ERDF A way of making Europe".

6. Conflict of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could inappropriately influence their work in this paper.

7. References

[1] Ballard, D., J. Winkler-Galicki, and J. Wesoly, *Massive parallel sequencing in forensics: advantages, issues, technicalities, and prospects.* Int J Legal Med, 2020. 134(4): p. 1291-1303.

[2] Stephens, K.M., et al., *Developmental validation of the ForenSeq MainstAY kit, MiSeq FGx sequencing system and ForenSeq Universal Analysis Software.* Forensic Sci Int Genet, 2023. 64: p. 102851.

[3] Ludeman, M.J., et al., *Developmental validation of GlobalFiler PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples.* Int J Legal Med, 2018. 132(6): p. 1555-1573.

[4] Bai, R., et al., *Developmental Validation of a novel 5 dye Y-STR System comprising the 27 YfilerPlus loci.* Sci Rep, 2016. 6: p. 29557.

[5] (SWGDAM), S.W.G.o.D.A.M., *Validation Guidelines for DNA Analysis Methods.* 2016: p. 1-15.

[6] Kling, D., A.O. Tillmar, and T. Egeland, *Familias 3 - Extensions and new functionality.* Forensic Sci Int Genet, 2014. 13: p. 121-7.

[7] Bleka, O., G. Storvik, and P. Gill, *EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts.* Forensic Sci Int Genet, 2016. 21: p. 35-44.

[8] Pestoni, C., et al., *Genetic data on three complex STRs (ACTBP2, D21S11 and HUMFIBRA/FGA) in the Galician population (NW Spain).* Int J Legal Med, 1999. 112(5): p. 337-9.

[9] Fernandez-Formoso, L., et al., *Allele frequencies of 20 STRs from Northwest Spain (Galicia).* Forensic Sci Int Genet, 2012. 6(5): p. e149-50.

[10] Barrio, P.A., et al., *Massively parallel sequence data of 31 autosomal STR loci from 496 Spanish individuals revealed concordance with CE-STR technology and enhanced discrimination power.* Forensic Sci Int Genet, 2019. 42: p. 49-55.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 899-904
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Massive sequencing in molecular autopsy. Gene panel vs whole exome sequencing

Alejandro Blanco-Verea[1], Estrella Sánchez-Carmona[1], Eva Ramos-Luis[1,2,3], María Álvarez-Barredo[3], Bernardo López-Abel[3,4], Rocio Gil[1], Ángel Carracedo[2,5], María Brion[1,2,3]

[1]*Cardiovascular Genetics, Santiago de Compostela Health Research Institute, Santiago de Compostela, Spain.*
[2]*Fundación Pública Galega de Medicina Xenómica (FPGMX), Santiago de Compostela, Spain.*
[3]*Inherited Cardiac Diseases Unit, Department of Cardiology, Santiago de Compostela University Hospital, Santiago de Compostela, Spain.*
[4]*Inherited Cardiac Diseases Unit, Department of Paediatric, Santiago de Compostela University Hospital, Santiago de Compostela, Spain.*
[5]*Institute of Forensic Sciences, Genomic Medicine Group-CIMUS, University of Santiago de Compostela, Spain.*

** Corresponding author at: María Brion, Xenética Cardiovascular, Instituto de Investigación Sanitaria de Santiago (IDIS), Spain, maria.brion@usc.es.*

Abstract

Molecular autopsy is presented as a fundamental tool to elucidate sudden unexplained death (SUD) cases. In the last decade, the detection of genetic variants has gone from Sanger sequencing to the most modern massive sequencing approaches allowing large gene panels to be analysed. The evolution of technology, together with the reduction in costs, is currently favoring the choice of whole exome sequencing (WES) as a molecular autopsy strategy. In this study we will try to evaluate the efficacy of re-evaluating by WES cases of SUD that were previously analysed for a panel of genes with inconclusive results.

In our study we started from 86 cases of sudden cardiac death (62 men and 24 women) with ages ranging from 0 to 62 years. All cases underwent a gene panel by massive sequencing that included the main genes associated with cardiomyopathies and channelopathies. This analysis showed pathogenic or likely pathogenic variants in relation to the risk of sudden cardiac death in 16% of the cases.

Among the non-conclusive cases, the sequencing study was extended to WES in 14 cases with structurally normal heart at autopsy or with mild findings that do not justify the diagnosis of

structural heart disease. Data analysis performed using a virtual panel of 565 genes related to cardiac diseases according to current evidence did not allow any additional variants related to sudden cardiac death to be identified.
Our study supports the efficacy of molecular autopsy in the search for the genetic cause of unexplained sudden cardiac death cases, showing a probable genetic cause in 16% of cases. However, it also suggests that the increase of genes beyond those definitively associated with cardiomyopathies and channelopathies, does not improve the efficacy of molecular autopsy. As long as there are no new genes with a definitive association to sudden cardiac death, new approaches should be directed towards the use of technologies that allow the detection of other genetic variants not detectable with WES, such as whole genome sequencing studies.

Keywords
Molecular autopsy, sudden unexplained death, massive sequencing, exome, gene panel.

1. Introduction

Molecular autopsy is presented as a fundamental tool to elucidate sudden unexplained death (SUD) cases. In the last decade, the detection of genetic variants has gone from Sanger sequencing to the most modern massive sequencing approaches allowing large gene panels to be analysed [1] in a short period of time and at a reduced cost.

The evolution of technology, together with the reduction in costs, is currently favoring the choice of whole exome sequencing (WES) as a molecular autopsy strategy [2] which provides access to more genetic information. In this study we will try to evaluate the efficacy of re-evaluating by WES cases of SUD that were previously analysed for a panel of genes with inconclusive results.

2. Material studied, methods, techniques

2.1. Material

In our study we started from 86 cases (62 men and 24 women) with ages ranging from 0 to 62 years, that were received in the laboratory for genetic study of sudden cardiac death (SCD) over the last 16 years.

This study was developed according to the recommendations of the Helsinki Declaration and approved by the local Ethical Committee. All cases or their relatives signed an approved written informed consent before enrollment.

2.2. Methods and techniques

Massive sequencing of a panel of the main genes associated with cardiomyopathies and channelopathies was performed in the 86 cases (Figure 1). Based on the evidence available over time, the number of genes included in the panel was variable, with a minimum of 81 and a maximum of 162 genes. Library preparation, sequencing technology and data analysis were also variable over time, according to technological development and availability in the laboratory.

A selection of 14 cases were subsequently subjected to whole exome sequencing with *KAPA HyperExome Probes (Roche*) and the *NovaSeq 6000 Sequencing System* from *Illumina*, and then prioritized to search for pathogenic variants using a virtual panel of 565 genes previously related with cardiac diseases (Figure1).

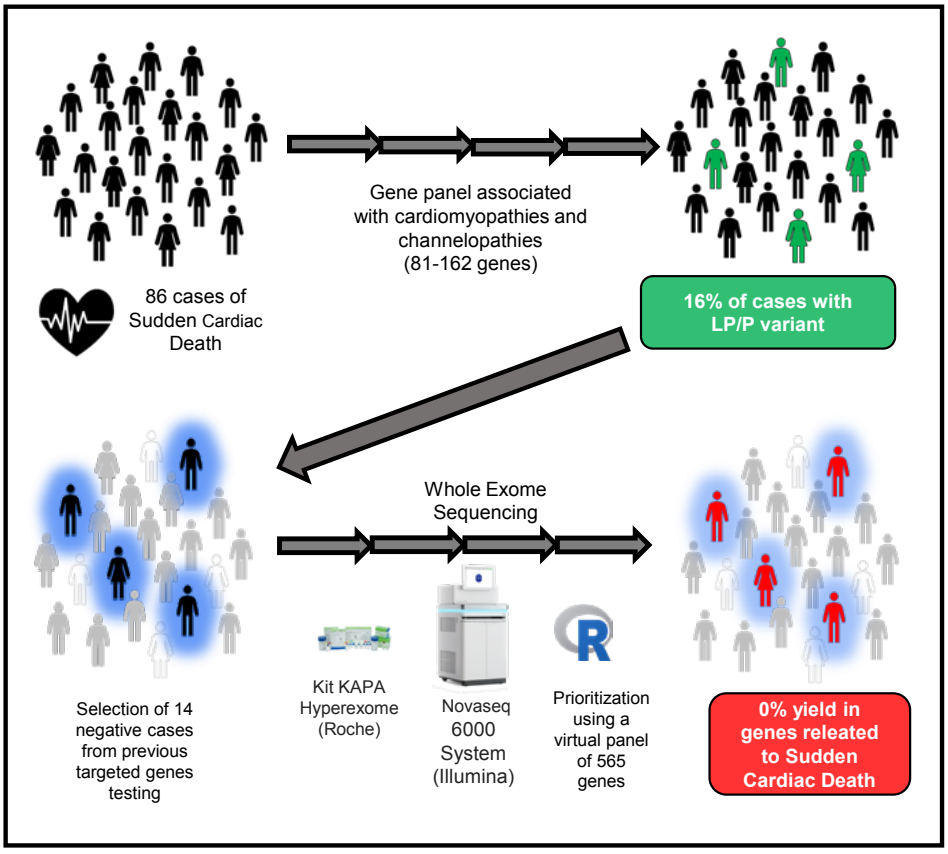


Figure 1. Steps to view the performance of WES versus specific gene panel in variants detection in SUD cases.

3. Results

Analysis of 86 cases with the inherited cardiac disease gene panel showed pathogenic or likely pathogenic variants in relation to the risk of sudden cardiac death in 14 cases (Table 1). The ACMG guidelines and their most recent updates were taken into account to determine the pathogenicity of the variants [3, 4]. In 6 of the 14 cases with putative causative variants, the autopsy showed structurally normal hearts, in 1 case the findings were compatible with a mild cardiac hypertrophy and the remaining 7 cases showed findings compatible with a previous undiagnosed cardiomyopathy.

Table 1. List of pathogenic or likely pathogenic variants found in 86 SUD cases sequenced with a gene panel. SID = Sudden Infant Death, DRV = Dilated Right Ventricle,

Cardiac findings at autopsy	Genetic variant
Structurally normal heart	*DSP*(NM_004415.2):c.1352G>A;p.R451H
Structurally normal heart	*KCNQ1*(NM_000218.2):c.727C>T;p.R243C
Structurally normal heart	*KCNH2*(NM_000238.3):c.1744C>T;p.R582C
Structurally normal heart	*KCNH2*(NM_000238.3):c.1772A>G;p.D591G
Structurally normal heart	*LMNA*(NM_170707.4):c.1583C>T;p.T528M
SID, Structurally normal heart	*MYH6*(NM_002471.3):c.2161del:R721Gfs*7
Mild cardiac hypertrophy	*RYR2*(NM_001035.2):c.1438C>T;p.R480*
Hypertrophic cardiomyopathy	*TNNT2*(NM_001001430):c.281G>A;p.R94H
Asymmetric biventricular hypertrophy	*NEXN*(NM_144573.3):c.784C>T;p.R262*
Biventricular hypertrophy	*MYBPC3*(NM_000256.3):c.1806del;p.I603Lfs*60
Asymmetric hypertrophy	*MYBPC3*(NM_000256.3):c.1806del;p.I603Lfs*60
Arrhythmogenic cardiomyopathy, severely DRV	*PKP2*(NM_004572.3):c.987del;p.S329Rfs*23
Dilated cardiomyopathy	*DES*(NM_001927):c.1109T>C;p.L370P
Arrhythmogenic left ventricular cardiomyopathy	*LMNA*(NM_170707):c.1364G>A;p.R455H

The genetic study expanded to a virtual panel of 564 genes analysed through whole exome sequencing, in a subsample of 14 cases, did not allow to identify new pathogenic or likely pathogenic variants for sudden cardiac death-related diseases.

4. Discussion

According to the current guidelines for management of ventricular arrhythmias and prevention of SCD [5], genetic testing is recommended when a condition with a likely genetic basis and a risk of ventricular arrhythmia and SCD is diagnosed in a living or deceased patient (class I). In addition, post-mortem genetic testing targeted to primary electrical diseases is recommended (class I) and may be extended to testing of additional genes (class IIb) in SCD victims younger than 50 years old or who had circumstances or family history supporting a primary genetic disease. Recommendations also include familial evaluation with genetic testing when post-mortem genetic testing detects a pathogenic mutation and the collection of blood samples for genetic testing after aborted SCD (class I).

Our study detected putative causative genetic variants in 16% of the cases, in accordance with recent data indicated at least a 13% genetic yield in SADS cases [5].

Since only 14 of the 72 cases inconclusive with gene panel analysis were subjected to an expanded genetic study, we cannot rule out that, in any of the other 58 cases, a variant associated with a cardiac disease could be detected with WES. It should also be taken into account that in the samples analysed by WES, a genetic putative causative variant may be present in another gene or genes not included in the virtual panel or even in intronic regions of the considered genes, as is the case of some pathogenic variants described in the MYBPC3 gene [6] associated with hypertrophic cardiomyopathy.

5. Conclusion

Our study supports the efficacy of molecular autopsy in the search for the genetic cause of unexplained sudden cardiac death cases, showing a probable genetic cause in 16% of cases. However, it also suggests that the increase of genes beyond those definitively associated with cardiomyopathies and channelopathies, does not improve the efficacy of molecular autopsy. As long as there are no new genes with a definitive association to SCD related diseases, new approaches should be directed towards the use of technologies, as whole genome sequencing, that allow the detection of other genetic variants in already known genes, not detectable with exonic sequencing approaches, such as intronic variants or structural variation.

6. Acknowledgments

Acknowledgements to Beatriz Sobrino, head of the Massive Sequencing Platform of the FPGMX.

Funding for this work was made possible thanks to funds from the *PI22/00286* project financed by the Instituto de Salud Carlos III, the *Plan de Recuperación, transformación y Resilencia* and *Next Generation EU* funds from the European Union.

7. Conflict of interest statement

No Conflict of interest.

8. References

[1] Castiglione V, Modena M, Aimo A, et al, Molecular Autopsy of Sudden Cardiac Death in the Genomics Era, Diagnostics (Basel). 11 (2021) 1378.

[2] Martínez-Barrios E, Grassi S, Brión M, et al, Molecular autopsy: Twenty years of post-mortem diagnosis in sudden cardiac death, Front Med (Lausanne). 10 (2023) 1118585.

[3] Richards S, Aziz N, Bale, et al, Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine. 17(5) (2015) 405-423.

[4] Hayesmoore JB, Bhuiyan ZA, Coviello DA, et al, EMQN: Recommendations for genetic testing in inherited cardiomyopathies and arrhythmias. European Journal of Human Genetics. 31(9) (2023) 1003-1009.

[5] Zeppenfeld K, Tfelt-Hansen J, de Riva M, et al, 2022 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death. European Heart Journal. 43 (2022) 3997-4126.

[6] Wood KA, Ellingford JM, Eden J, et al, Pathogenic intronic Splice-Affecting Variants in MYBPC3 in Three Patients with Hypertrophic Cardiomyopathy. Cardiogenetics. 11 (2021) 73-83.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 905-911
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Systemic approach: A review of forensic genetic programs working on Missing Person Identification in humanitarian contexts

Yarimar Cristina Ruiz Orozco[1*]

[1]*International Committee of the Red Cross (ICRC), Geneva, Switzerland.*

** Avenue de la Paix 19, Geneva 1202, Switzerland, +41227302816, yruizorozco@icrc.org*

Abstract

The Identification of Missing Persons (MPI) in communities affected by armed conflicts and other situations of violence is a fundamental pillar to address humanitarian needs. Furthermore, the integration of forensic genetic programs into the MPI has become an added value to this multidisciplinary process. The approach of strengthening medicolegal systems - the systemic approach - has been described as an effective strategy for the resolution of MPI. This article reviews under the systemic approach three forensic genetic programs working in MPI based twenty-eight indicators clustered in seven requirements. Common challenges were identified in the three humanitarian contexts, mainly related to implementation of policies and the importance to address Family's needs, impacting the resolution of MPI. The systemic approach enables the identification of gaps and the integration of needs when assessing the functioning of forensic genetic programs working in MPI.

Keywords

Missing Person Identification (MPI), Humanitarian Contexts.

1. Introduction

The Identification of Missing Persons (MPI) in communities affected by armed conflicts and other situations of violence is a fundamental pillar to address humanitarian needs. Furthermore, the integration of forensic genetic programs into the MPI process has become a valuable addition to this multidisciplinary approach.

Several recommendations have been published on the use of forensic genetics to address the issue of MPI in the last two decades [1,2,3,4]. There is still,

however, a need to carry out an integrated process leading to the clarification of the fate and whereabouts of the missing, which is a much complex reality – especially in disasters, migration scenarios, armed conflict, and other situations of violence, where it is not possible to assume in the first instance that the person is alive or dead and where a coordinated, integrated, multidisciplinary and very often multi-agency response is required [5].

Strengthening medicolegal systems - a systemic approach - has been described as an effective strategy for resolving MPI cases by M.D Morcillo in 2024 [6], providing a general overview of the main gaps and challenges of the system that prevent States from responding adequately, through a set of recommendations to adjust the development and implementation of programs working on MPI. This study reviews, under the systemic approach, three cases of forensic genetic programs working on MPI in humanitarian contexts.

2. Material and Methods

The three genetic programs working under humanitarian contexts[1] were assessed in the current study as follows:

- Context 1: Other Situation of Violence (OSV), represented by seven laboratories working under forensic services.
- Context 2: Non-International Armed Conflict (NIAC), represented by one laboratory working under a medico-legal directorate.
- Context 3: International Armed Conflict (IAC), represented by one laboratory working under a mechanism of missing persons.

The seven requirements described by Morcillo (2024) for the proper functioning of medicolegal systems - policy, procedures, human resources, infrastructure, information management, quality, and family engagement - were implemented following standardized assessment [7, 8] for the three forensic genetic programs focused on MPI in humanitarian contexts. To measure each of the seven requirements, four indicators per requirement were adapted for this study, totaling twenty-eight indicators as follows:

- Policy: Laws, regulations developed and implemented in aspects related to i) Missing Person Identification, ii) DNA use, iii) DNA databases and,

[1] Name of countries and institutions are been anonymized

iv) national or international committees or mechanism for the search and identification of missing persons.
- Procedures: development and implementation of procedures and protocols related to i) Inter institutional cooperation, ii) intra institutional cooperation, iii) Methods, SOPs, iv) Integrated identification process (multidisciplinary approach).
- Families: Implementation of aspects related to the participation of families in the search and identification process i) Missing Person Data collected, ii) Biological References Samples collected, iii) Information centers available (information shared), iv) accompaniment of families in the process (participation).
- Infrastructure: working conditions related to i) dedicated structures, ii) adequate distribution of spaces, iii) adequate equipment, iv) optimization and automation of process.
- Human Resources: working conditions related to i) funding for projects related, ii) dedicated organigram and job descriptions (HR set up), iii) skills on forensic genetics and human identification, iv) academic curricula on forensic genetic and human identification.
- Information Management: proper collection, creation, processing, analysis, sharing, use, archiving/destruction, delivery of data related to i) development of databases, ii) databanks, iii) data management system in place, iii) data protection practices in place.
- Quality: implementation of aspects related to i) practices of quality assurance, ii) practices of quality control, iii) corrective and preventive maintenance plan in place, iv) accreditation process on going or done and maintained.

3. Results

The evaluation of the twenty-eight indicators were represented in a Presence-Absence Matrix (PAM) to compare the three forensic genetic programs under the humanitarian contexts of OSV, NIAC and IAC (Figure 1). Seven indicators were met in all the genetic programs assessed: National or International committees or mechanism for the search of Missing Persons; Methods and SOPs*; Dedicated organigram and job description (HR set up); Dedicated structures,* Biological Reference Samples (BRS) collected*;* Development of databases*; and* Quality Control. In the other hand, eleven indicators were not met in any of the three genetic programs assessed: Integrated identification process (multidisciplinary); Funding for projects

related; Skills on forensic genetic and human identification; Optimization and automation of processes; Information Centre available (information shared); Accompaniment of Families in the process (participation); Development of databanks; Data Management System; Quality assurance; Corrective and preventive maintenance; and Accreditation process. The frequency of overall indicators met in each of the three evaluated genetic programs were represented in a Treemap chart (Figure 2).

It was also evaluated the three genetic programs working on MPI based on the seven requirements described by M.D Morcillo, represented in a heat map chart (Figure 3). Notably, a higher frequency of indicators was observed for Policies (4/4) compared to a lower frequency for Quality (1/4). Finally, when comparing the functioning of the three forensic genetic programs, a higher rate of indicators was met for the program under the context of OSV at 57%, followed by NIAC at 39% with the lowest rated observed for IAC at 32% as shown in the funnel chart (Figure 4).

4. Discussion

Common challenges in the functioning of forensic genetic programs working on MPI were identified across the three humanitarian contexts, mostly coinciding on the need to implement policies and regulations, which affects other requirements essential to addressing families' needs. This analysis through the systemic approach, also reveals how the interconnection between indicators from different requirements further impacts the overall effectiveness of the programs. For instance, missing person data may be collected (Families) but not integrated in the identification process (Procedures); equipment may be in place (Infrastructure) without a maintenance plan (Quality); or database regulation may exist (Policy) without a corresponding data management system (Information Management).

This analysis highlights the importance of assessing case resolution of missing persons by integrating the measurement of elements from different nature (legal, scientific, social, cultural) through the development of indicators which are interconnected within each system.

5. Conclusion

The functioning of forensic genetic programs in MPI requires to be integrated, understood, and discussed within the framework of a systemic approach, which enables better identification of gaps, assessment of sustainable impact and provision of effective prevention and resolution of cases.

6. Acknowledgments

The author would like to thanks to the Forensic Unit of the ICRC, specially to Pierre Guyomarc'h, Jane Tylor and Claudia Garrido for their comments to the poster presented at ISFG 2024 congress. Likewise, the author would like to thanks to the iDevelop program of the ICRC for supporting participation to the ISFG 2024 congress.

7. Conflict of interest statement

No conflict of interest.

8. References

[1] J. Ward, Best practice recommendations for the establishment of a national DNA identification program for missing persons: A global perspective, FSI; Genetics Supplement Series 2 (2017) e43-e45.

[2] INTERPOL DNA Monitoring Expert Group, Best practice principles: Recommendations on the use of DNA for the identification of Missing Persons and unidentified human remains (2015)

[3] Scientific Working Group on DNA Analysis Methods (SWGDAM), Guidelines for Missing Persons casework (2014)

[4] M. Prinz, A. Carracedo, W.R Mayr, et al. DNA commission of the International Society of Forensic Genetics (ISFG): Recommendations regarding the role of forensic genetics for disaster victim identification (DVI), FSI: Genetics (2007) 3-12

[5] M. Salado-Puerto, D. Abboud, J. P. Baraybar et al,. The search process: Integrating the investigation and identification of missing and unidentified persons, FSI Synergy (2021) 3: 100154

[6] M.D. Morcillo-Mendez, Strengthening the medicolegal system: Fulfilling international law obligations during conflicts and disasters to prevent and resolve issues of humanitarian concern, International Review of the Red Cross, Page 1 of 28. doi.org/10.1017/S181638312400002X

[7] ICRC, Guidelines for the use of forensic genetics in investigations into human rights and international humanitarian law violations, Ministerio de Relaciones Exteriores y Cultura, República Argentina (2019)

[8] ICRC, Medico-legal systems and forensic services: an assessment guide, Central Tracing Agency (2022)

	Indicator	IAC	NIAC	OSV
Policy	Missing Person Identification	1	1	1
	DNA use	0	1	1
	DNA databases	0	0	1
	Nat. or Int. committees or mechanism for the search of MP	1	1	1
Procedures	Inter-institutional cooperation	0	0	1
	Intra-institutional cooperation	0	1	1
	Methods and SOPs	1	1	1
	Integrated identification process (multidisciplinary)	0	0	0
Families	Missing persons data collected	0	1	1
	Biological Reference Samples (BRS) collected	1	1	1
	Information Centre available (information shared)	0	0	0
	Accompaniment of Families in the process (participation)	0	0	0
Infrastructure	Dedicated structures	1	1	1
	Adequate distribution of spaces	1	0	0
	Adequate equipment	0	0	1
	Optimization and automation of processes	0	0	0
Human Resources	Funding for projects related	0	0	0
	Dedicated organigram and job description (HR set up)	1	1	1
	Skills on forensic genetic and human identification	0	0	0
	Academic curricula on forensic genetic and human ID	0	0	1
Information Management	Development of databases	1	1	1
	Development of databanks	0	0	0
	Data management system	0	0	0
	Data protection practices	0	0	1
Quality	Quality assurance	0	0	0
	Quality Control	1	1	1
	Corrective and preventive maintenance	0	0	0
	Accreditation process	0	0	0
	Total	9	11	16

Figure 1. Presence Absence Matrix (PAM) of the twenty-eight indicators for comparing the three forensic genetic programs working under humanitarian contexts.

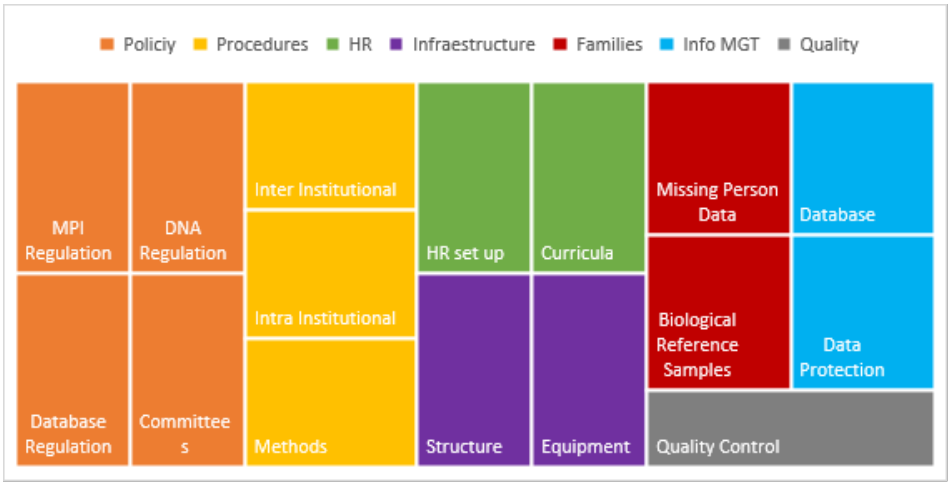


Figure 2. Treemap chart frequency of indicators observed in the contexts evaluated.

Context	Policy	Procedures	Families	Infrastructure	Human Resources	Information Management	Quality
IAC	2	1	1	2	1	1	1
NIAC	3	2	2	1	1	1	1
OSV	4	3	2	2	2	2	1

Figure 3. Heatmap chart of the evaluation of each requirement assessed per forensic genetic program under three humanitarian contexts

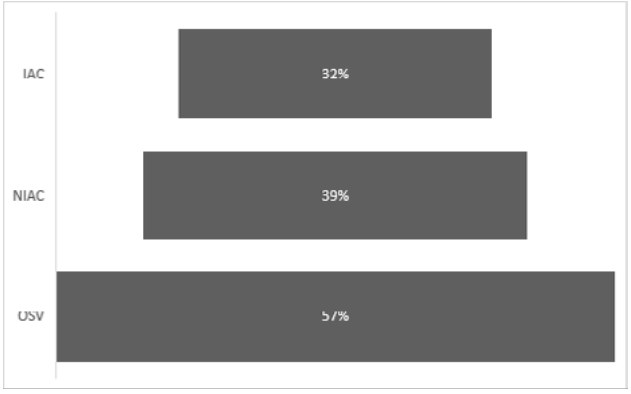


Figure 4. Funnel chart evaluation (in percentage) of each context based on the overall indicators assessed

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 913-919
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Enhancing minor component detection in unbalanced mixtures using Multi-Indel

Mengna Wu[1], Jinlong Song[1], Yazi Zheng[1], Guihong Liu[1], Qiushuo Wu[1], Xiameng Chen[1], Lin Zhang[1], Shengqiu Qu[1*], Weibo Liang[1*]

[1]*Department of Forensic Genetics, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu, 610041, Sichuan, China*

** Corresponding author at: Fax number: +86 28 85401825; E-mail address: qushengqiu@scu.edu.cn (S. Qu), liangweibo@scu.edu.cn (W. Liang)*

Abstract
Mixtures have always been a challenging problem in forensic genetics. Traditional genetic markers, like short tandem repeat (STR), can typically address ratios of up to 1:20, but when there is a significant difference between individual components, especially in extremely unbalanced mixtures (EUM), STRs struggle to provide reliable results. We propose designing allele-specific primers for Multi-Indels to target the minor component within mixtures. This method allows for the specific amplification of the minor component in cases where the major and minor components are homozygous for opposite alleles, or when the major component is homozygous and the minor component is heterozygous. Ultimately, we screened 11 Multi-Indels from the 1000 Genomes Project, each achieving a sensitivity of 10 pg and successfully detecting the minor component in a 1:1000 mixture.

Keywords
Multi-indel, allele-specific amplification, unbalanced mixture.

1. Introduction

Mixtures are very common in forensic science, particularly in cases involving two components. The greater the imbalance of proportions, the more challenging the analysis becomes, a phenomenon known as EUM. PCR amplification exacerbates this imbalance, potentially leading to the loss of alleles in subsequent genotype

profiles, an issue that traditional genetic markers struggle to address. In recent years, some studies have reported the use of compound genetic marker to separate and detect the mixture based on the allele-specific amplification methods, including deletion/insertion polymorphism-STR combinations (DIP-STR) [1], single nucleotide polymorphism-STR combinations (SNP-STR) [2], DIP-microhaplotypes [3], DIP-SNP [4], and SNP-SNP [5]. Each compound marker has its pros and cons. DIP-STR is sparsely distributed across the genome and is difficult to match with existing STR. SNP-STR requires the artificial introduction mismatches at the 3' end of the allele-specific primer, which can be hard to control and may increase non-specific amplification products. DIP-microhaplotypes, DIP-SNP and SNP-SNP necessitate combing allele-specific amplification with SNapshot assay, complicating the process due to the need for two reactions.

Multi-indels, a generalized microhaplotype formed by two or more closely linked indels within 200-300bp, has emerged as a promising candidate [6]. Compared to other compound genetic markers, their advantage lies in widespread distribution across the genome. Additionally, Multi-indels exhibit length polymorphism, making it compatible with capillary electrophoresis (CE) platforms without the need for introducing mismatches or other techniques. Consequently, we screened Multi-indels from the whole genome and conducted allele-specific amplification.

2. Material and methods

2.1. Selection of Multi-indel markers

We used VCFtools to select Multi-Indels from the 103 Han Chinese in Beijing (CHB) and 105 Southern Han Chinese (CHS) of the 1000 Genomes Project Phase 3. The selection criteria were as follows: (1) For indels: being biallelic; minor allele frequency (MAF) > 0.1; allele length ranging from 1 to 30 bp; located in introns or non-coding regions. (2) For Multi-Indels: Each Multi-indel consists of at least 3 indels, with a total length within 200 bp; excluding the first or last indel, the length of any indel must not be equal to the length of any other indel or the sum of two or more indels, ensuring that each theoretical haplotype has a unique amplicon length; no other indels or highly repetitive nucleotide sequences within the amplicon; the physical distance between different Multi-Indels on the same chromosome must be >10 Mb; the theoretical haplotype frequency calculated by Phase must be ≥4, with at least 3 haplotypes having a frequency ≥0.05.

2.2. Primer design

For each Multi-Indel, we designed allele-specific primers targeting the different alleles of the first indel, referred to as long and short alleles, with corresponding primers labeled as L-primer and S-primer. During PCR, these primers specifically amplify the long and short sequences. The high specificity is achieved through the difference between the insertion sequence in the L-primer and the sequence spanning the insertion site in the S-primer. Additionally, a shared reverse primer (R-primer) was designed for the sequence downstream of the last indel. By controlling the amplicon lengths of different alleles, individual haplotypes can be directly determined on CE platforms. To further enhance specificity, the L and S primers were used in separate reactions (L-reaction and S-reaction).

2.3. Sensitivity study

The sensitivity of each Multi-Indel was assessed using samples with heterozygous indels for which allele-specific primers were designed. These samples were diluted to 50 pg, 20 pg, and 10 pg for singleplex amplifications of L and S reactions in triplicate.

2.4. Mixture study

We assessed the ability to detect the minor component in mixtures of each Multi-indel. For the L and S primers of each locus, the simulated mixtures consisted of heterozygous samples and corresponding homozygous samples. LS+SS was used for detection with L primers, while LS+LL was used for detection with S primers. The selection of heterozygous and homozygous samples depended on the allele of the indels used to design the allele-specific primers. Based on the sensitivity experiments results, the minor component was fixed at 10 pg, while the major components were set at 1 ng, 2 ng, 5 ng, and 10 ng, corresponding to mixture ratios of 1:100, 1:200, 1:500, and 1:1000.

3. Results

We screened a total of 629,402 indels from the 1000 Genomes Project data, meeting the criteria of being biallelic, with MAF > 0.1 and allele length less than 30 bp. Next, we filtered for Multi-Indels that contained at least 3 indels, with a physical distance within 200 bp, ensuring each haplotype had a unique length. This process resulted in 2,860 candidates, all located in introns or non-coding regions. Based on

the number and frequency of haplotypes calculated by Phase software, we required a theoretical haplotype count ≥ 4 with at least 3 haplotypes having a frequency ≥ 0.05, yielding 2,580 candidates. We then excluded loci with nucleotide repeats and indels overlapping, ensuring no other length variance within the amplicon and compliance with primer design standards, resulting in 59 candidates. Finally, we retained the Multi-indels with good primer specificity and no obvious non-specific products, excluding those on the same chromosome that were less than 10 Mb apart, ultimately remaining 11 Multi-Indels. General information of these 11 Multi-indels is provided in Table 1.

Regarding sensitivity and the ability of detect unbalanced mixture, all primers achieved a sensitivity of 10 pg and were capable of detecting simulated mixtures with a ratio of 1:1000. Here we present the electrophoresis results for mh05ZL-008. Figure 1 and 2 display the sensitivity and mixture results for the L and S primer of mh05ZL-008. The left half of each figure shows the results for the L-primer, while the right half presents the results for the S-primer. The results are arranged from top to bottom for 10 pg, 20 pg, and 50 pg, as well as for mixture ratios of 1:100, 1:200, 1:500, and 1:1000, respectively. The genotype of the heterozygous sample used for sensitivity assay is mh05ZL-008-L-0 and mh05ZL-008-S-3. The genotype of minor DNA in mixture is mh05ZL-008-L-0 and mh05ZL-008-S-3, while the genotypes of major DNA are mh05ZL-008-S-3 for the L-primer and mh05ZL-008-L-0 for the S-primer. From the figures, it is evident that mh05ZL-008 achieves a sensitivity of 10 pg and can detect minor components from a 1:1000 extremely unbalanced mixture.

4. Discussion and conclusion

We screened 11 Multi-indels suitable for detecting extremely unbalanced mixtures using allele-specific amplification technology from the 1000 Genomes Project data. These Multi-indels can achieve a sensitivity of 10 pg and detect minor components in simulated mixtures up to 1:1000. However, our study has some drawbacks, the number of loci is relatively small, which may not suffice for individual identification or parentage testing, serving primarily as a supplement to STRs. Additionally, the requirement for two reactions to detect individual genotypes increases the DNA quantity needed.

5. Acknowledgments

This work was supported by grants from China Postdoctoral Science Foundation (2022M722246).

6. Conflict of interest statement

None.

7. References

[1] V. Castella, J. Gervaix, D. Hall, DIP-STR: highly sensitive markers for the analysis of unbalanced genomic mixtures, Hum Mutat 34(4) (2013) 644-54.

[2] Y. Tan, P. Bai, L. Wang, H. Wang, H. Tian, H. Jian, R. Zhang, Y. Liu, W. Liang, L. Zhang, Two-person DNA mixture interpretation based on a novel set of SNP-STR markers, Forensic Science International: Genetics 37 (2018) 37-45.

[3] J. Liu, T. Hao, X. Cheng, J. Wang, W. Li, Z. Liu, J. Shi, Z. Li, J. Ren, K. Yun, G. Zhang, DIP-microhaplotypes: new markers for detection of unbalanced DNA mixtures, International journal of legal medicine 135(1) (2020) 13-21.

[4] Z. Liu, J. Liu, J. Wang, D. Chen, Z. Liu, J. Shi, Z. Li, W. Li, G. Zhang, B. Du, A set of 14 DIP-SNP markers to detect unbalanced DNA mixtures, Biochemical and Biophysical Research Communications 497(2) (2018) 591-596.

[5] R. Zhang, Y. Tan, H. Jian, S. Qu, Y. Liu, J. Zhu, L. Wang, M. Lv, M. Liao, L. Zhang, F. Yang, W. Liang, A new approach to detect a set of SNP-SNP markers: Combining ARMS-PCR with SNaPshot technology, Electrophoresis 41(13-14) (2020) 1189-1197.

[6] S. Qu, M. Lv, J. Xue, J. Zhu, L. Wang, H. Jian, Y. Liu, R. Zhang, L. Zha, W. Liang, L. Zhang, Multi-Indel: A Microhaplotype Marker Can Be Typed Using Capillary Electrophoresis Platforms, Frontiers in genetics 11 (2020) 567082.

Table 1. General information of 11 Multi-indels

Multi-indel	Chromosome	Indel allele	Rs number	Position (GRCh37/hg19)
mh02ZL-021	2	-/CACTT	rs71933401	10929385
	2	-/AT	rs11334495	10929488
	2	-/GAAAAC	rs57608395	10929533
mh04ZL-012	4	-/TC	rs35846816	88801445
	4	-/GAACT	rs72161488	88801519
	4	-/CCCTT	rs34948283	88801550
mh04ZL-013	4	-/CAT	rs66502037	187124231
	4	-/GCACA	rs77222977	187124238
	4	-/TTCATA	rs71871946	187124313
mh05ZL-007	5	-/CAGGTGAAAAATTAG	rs113305200	103101444
	5	-/TAGAGAA	rs375318992	103101611
	5	-/ATTTCCGTT	rs369927246	103101622
mh05ZL-008	5	-/TA	rs33954215	164476543
	5	-/AATT	rs33931627	164476563
	5	-/CCATT	rs80340616	164476631
mh06ZL-012	6	-/AGTCACTATT	rs144153898	32560182
	6	-/CAGG	rs148954634	32560270
	6	-/AT	rs537623658	32560304
mh07ZL-014	7	-/CTAAATGAT	rs71053237	57322877
	7	-/TATA	rs72447238	57322974
	7	-/AAT	rs71053238	57322980
mh14ZL-006	14	-/TAAAACACTG	rs76021271	37149167
	14	-/TATC	rs5807900	37149210
	14	-/AAAAT	rs144923203	37149247
mh18ZL-011	18	-/AGAGGTGGGACC	rs377195018	61672654
	18	-/TTGGG	rs59925455	61672667
	18	-/GAT	rs201823781	61672670
mh19ZL-006	19	-/AGGCCTGGCGC	rs140364268	48264590
	19	-/GTT	rs140784512	48264662
	19	-/CGGCTA	rs60008794	48264754
mh21ZL-004	21	-/GA	rs11424626	43550540
	21	-/GTAT	rs76728303	43550582
	21	-/TGTGA	rs10671342	43550724

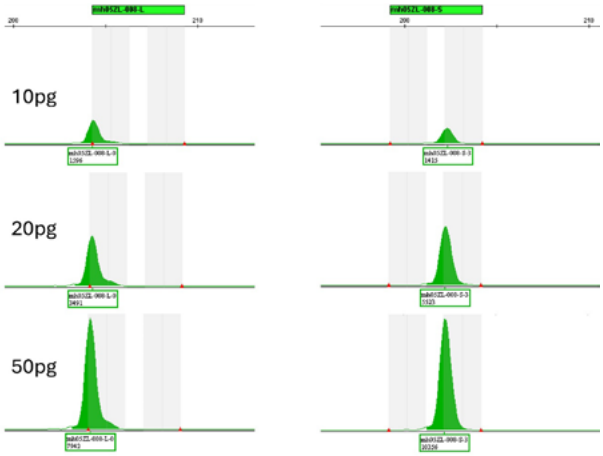


Figure 1. The sensitivity results of the L and S primer for mh05ZL-008.

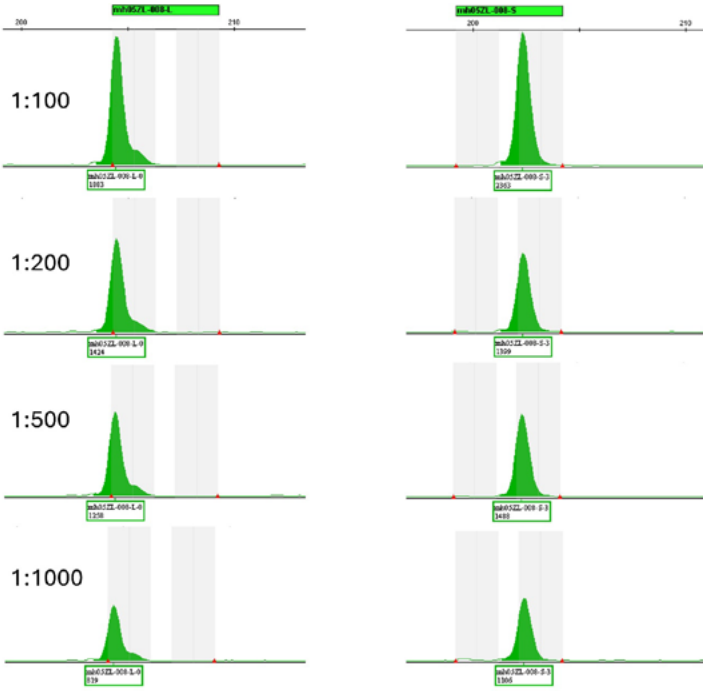


Figure 2. The mixture results of the L and S primer for mh05ZL-008.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 921-928
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Plasma exosome-associated piRNA screening in acute myocardial infarction based on high-throughput sequencing

Jia Zhang[a,b,1], Peng Zhou[a,b,1], Ruotong Xiong[a,b], Yujie Wang[a,b], Min Li[a,b], Deping Meng[a,b], Xiaoqun Xu[a,b], Chunjiang Yu[a,b], Ran Wei[a,b], Jiangwei Yan[b,c*], Chen Fang[a,b*]

[a]*Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan 250021, P. R. China;*

[b]*Department of Forensic Medicine, School of Clinical and Basic Medical Sciences, Shandong First Medical, University& Shandong Academy of Medical Sciences, Jinan 250117, P. R. China;*
[c]*Shanxi Medical University, Taiyuan 030001, P. R. China.*

** Corresponding author at:*
Dr Jiangwei Yan, Shanxi Medical University, Taiyuan 030001, P. R. China.
Email: yanjw@sxmu.edu.cn (Jiangwei Yan)
Dr Chen Fang, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan 250021, P. R. China.
Email: chenfangyulix@126.com (Chen Fang)
1 These authors contributed equally to this work.

Abstract

Objective: Acute myocardial infarction (AMI) is an important cause of sudden cardiac death, and the accurate and rapid diagnostic methods for AMI remain insufficiently developed. PIWI-interacting RNAs (piRNA), as a newly discovered type of small RNA, have been found to be associated with the occurrence of cardiovascular diseases (CVD), but fewer studies have been conducted on the association of piRNAs with AMI. In this study, the piRNAs in plasma exosomes that are associated with the development of AMI was explored based on high-throughput sequencing technology, and preliminarily investigated the correlation between piRNAs and AMI.

Methods: An AMI model was established in rats by ligating the anterior descending branch of the left coronary artery. The rats were euthanized at different time points following the establishment of the AMI model. RNA was extracted from rat plasma exosomes and high-throughput sequencing technology was utilized to screen for differentially expressed piRNAs in rat plasma exosomes associated with AMI.

Results: A total of 109,947 piRNAs were detected, 212 piRNAs showed an upward trend and 88 piRNAs showed a downward trend in expression with the prolongation of AMI. Target gene

analysis of the top few piRNAs with more significant up- and down-regulation associated with AMI revealed that, for example, piR-2490770 and piR-282889 have the same target gene, Mterf3, which is involved in negative transcriptional regulation.
Conclusion: The expression of piRNA in plasma exosomes changes at the onset of AMI. This study contributes to the discovery of novel molecular markers suitable for the early diagnosis of AMI and provides new ideas for recognizing and preventing AMI.

Keywords
acute myocardial infarction; piRNA; plasma exosome; high-throughput sequencing

1. Introduction

Acute myocardial infarction (AMI) is one of the most serious cardiovascular diseases(CVD), characterized by high rates of morbidity and mortality[1]. Commonly used clinical methods of AMI identification, such as imaging methods that have limited significance in the post-mortem diagnosis of early AMI without vascularized organic lesions, have not yet been universally applied to AMI post-mortem identification[2]. Therefore, it is crucial to identify new biomarkers that can enhance the accuracy of AMI diagnosis.

Exosomes, which facilitate cellular communication, are widely involved in various biological processes such as immune response, cell differentiation, tumor growth[3]. Studies have shown that exosomes are involved in cardiovascular pathophysiological processes[4]. Additionally, piwi-interacting RNAs (piRNAs) have been identified as important regulatory elements in CVD conditions such as heart failure, cardiac hypertrophy, pulmonary hypertension, and ischemic cardiomyopathy[5]. PiRNAs are involved in silencing transcriptional genes, maintaining germline and stem cell functions, regulating translation and mRNA stability, maintaining developmental and genetic integrity[6].

The expression of RNA in the body is temporal and spatial, and the amount of expression will change at different stages of its generation process or under the influence of factors such as disease[7]. Currently, there are fewer studies on AMI-related piRNAs. In this study, we employed high-throughput sequencing technology to identify AMI-related piRNAs. This approach not only offers new biomarkers for the post-mortem differential diagnosis of AMI in forensic medicine but also opens up further possibilities for prognostic treatment of AMI.

2. Material studied, methods, techniques

2.1. Construction of rat AMI model

Fifteen adult male Wistar rats were allocated into three groups: Sham, AMI 30-minute, and AMI 60-minute. Anesthesia was induced intraperitoneally using 10% chloral hydrate. The left anterior descending coronary artery was ligated, and the duration of ligation was categorized as 30 minutes for the AMI 30-minute group and 60 minutes for the AMI 60-minute group. In the Sham group, a thread was passed around the artery but not tied. The physiological status of the rats in all three groups was monitored through electrocardiographic (ECG) changes following the surgical procedure

2.2. Blood plasma isolation and exosome extraction

Supernatant was further centrifuged at 4°C at 3000×g for 15 minutes to isolate the plasma. Exosomes were extracted using the Hieff® Quick Exosome Isolation Kit (for Serum/Plasma) (Yeasen Biotechnology, Shanghai, China). The morphological structure of the exosomes was examined using transmission electron microscopy (TEM). Additionally, the particle size and distribution of the exosomes were analyzed using nanoparticle tracking analysis (NTA), and the presence of exosomal marker proteins (TSG101, CD81, Flotillin-1) was confirmed through Western blotting.

2.3. RNA extraction and high-throughput sequencing

RNA was extracted from exosomes utilizing the miRcute miRNA isolation kit (Tiangen Biotech, Beijing, China). The sequencing reads were alignmented with the piRBase databases [8]. Trend analysis was conducted using the Short Time-series Expression Miner (STEM) software [9]. Furthermore, the target genes associated with the identified AMI-related piRNAs were retrieved from the piRBase database.

2.4. Statistical analysies

GraphPad Prism 9 (GraphPad Software, La Jolla, CA, USA) statistically analyzed piRNAs screened by high-throughput sequencing technology with P value < 0.05 and multiplicity of difference in TPM > 2.

3. Results

3.1. Construction of rat AMI model

The S-T segment elevation of ECG in AMI 30min and AMI 60min groups suggested that the rat AMI model was successfully constructed.

3.2. Exosome extraction and identification

The exosome characteristic protein was verified to be positive by Western blot. The structure of extracellular vesicles was a double-layered vesicular spherical from TEM. The size range of the exosome was 50-150 nm, and the main peak was about 80 nm from NTA. In conclusion, the exosome extraction was successful.

3.3. Sequencing and data Analysis

A total of 109,947 piRNAs were detected. There were 15,634 up-regulated piRNAs and 9,100 down-regulated piRNAs in the AMI 60min group compared to the Sham group. There were 8088 up-regulated piRNAs and 10909 down-regulated piRNAs in the AMI 30min group compared to the Sham group. Two hundred and twelve piRNAs show an upward trend and eighty-eight piRNAs show an downward trend in expression after AMI occured. The top 15 piRNAs with significant upward and downward trends were shown in Figure 1.

3.4. Target gene analysis

Fifteen piRNAs each with more significant up-regulation and down-regulation from the previous screening were analyzed for target genes, and a total of seven target genes were screened(Table 1). Two of them, piR-rno-2490770 and piR-rno-282889, had the same target gene. Mterf3 is also involved in negative transcriptional regulation as well as DNA templating and activation of transcriptional cis-regulatory region binding activity in humans. Wwp2 is associated with protein ubiquitination. In addition, it has been found that miR-140-5p exacerbates hypoxia-induced cellular injury at the onset of infarction in part by upregulating MLK3[10].

4. Discussion

AMI is one of the most serious diseases in CVD, with a sudden onset that can lead to severe outcomes, including death. The continuous improvement and discovery of novel methods for the differential diagnosis of AMI hold great significance in forensic medicine.

In this study, a total of 300 piRNAs associated with AMI were identified through sequencing, and the expression of 212 piRNAs was positively correlated with the duration of AMI occurrence, while 88 piRNAs were negatively correlated. Compared to the 195 up-regulated and 13 down-regulated piRNAs found in the

serum of AMI patients reported by Ying Huang et al. [11], the number of piRNAs identified in this study is relatively large.

Previous studies have shown that piRNAs are involved in many diseases, such as cancer and Parkinson's disease. They have been found to play an important regulatory role in cardiovascular diseases (CVD) like heart failure and ischemic cardiomyopathy. Serena Vella et al. discovered that piRNAs were associated with reduced ischemic damage to the heart[14]. The expression of piR-DQ593039 in extracellular vesicles is upregulated in patients with pulmonary arterial hypertension, which may serve as a potential biomarker for lung and heart diseases[15]. Moreover, the expression of piR-5938 was increased during cardiac ischemia and reperfusion, and its targeted binding to miR-324 and miR-668 inhibited cardiomyocyte apoptosis[16]. It can be observed that piRNAs, similar to miRNAs, represent a class of small molecule RNAs that are stable and sensitive in response to myocardial injury. Therefore, piRNAs have the potential to be used as biomarkers for identifying AMI.

For the AMI-related piRNAs screened from rat plasma exosomes, which have not been reported in related CVD studies, the target genes and biological functions of some of the piRNAs have yet to be elucidated and require further exploration. This study offers new ideas for forensic AMI identification and expands the application of piRNAs in forensic science.

5. Conclusion

In the present study, piRNA expression profiles in plasma exosomes at the onset of AMI were analyzed based on high-throughput sequencing technology.We identified piRNAs with altered expression levels that may serve as potential biomarkers for AMI diagnosis. These findings also offer novel insights for forensic identification of AMI.

6. Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 82002006 and 82030058), Qingchuang Talents Induction Program of Shandong Higher Education Institution (2022 Forensic medicine innovation team), Key Research and Development Program of Shandong Province (Major Science and Technology Innovation Project) Project (2021CXGC011305).

7. Conflict of interest statement

The authors declare that they have no conflict of interest.

8. References

[1] Mechanic OJ, Gavin M, Grossman SA, Acute Myocardial Infarction, [Updated 2023 Sep 3]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-.

[2] Dong T N, Li N, Wang X X, et al. Research Progress of Forensic Diagnosis Approaches of Early Acute Myocardial Infarction, J. Fa yi xue za zhi, 2022, 38(3): 385-395.

[3] Yi J U, Ali S A H, Getao S H I, et al. Research Progress of Exosomes as Drug Delivery Carriers, J. Progress in Pharmaceutical Sciences. 2023. 47(11): 804-816.

[4] Han C, Yang J, Sun J, et al. Extracellular vesicles in cardiovascular disease: biological functions and therapeutic implications, J. Pharmacology & therapeutics, 2022, 233: 108025.

[5] Li M, Yang Y, Wang Z, et al. Piwi-interacting RNAs (piRNAs) as potential biomarkers and therapeutic targets for cardiovascular diseases, J. Angiogenesis, 2021, 24: 19-34.

[6] Li B, Wang K, Cheng W, et al. Recent advances of PIWI-interacting RNA in cardiovascular diseases, J. Clinical and Translational Medicine, 2024,s 14(8): e1770.

[7] Huang X, Yuan T, Tschannen M, et al. Characterization of human plasma-derived exosomal RNAs by deep sequencing, J. BMC genomics, 2013, 14: 1-14.

[8] Wang J, Zhang P, Lu Y, et al. piRBase: a comprehensive database of piRNA sequences, J. Nucleic acids research, 2019, 47(D1): D175-D180.

[9] Ernst J, Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data, J. BMC bioinformatics, 2006, 7: 1-11.

[10] Xing B, Li Q J, Li H, et al. miR-140-5p aggravates hypoxia-induced cell injury via regulating MLK3 in H9c2 cells[J]. Biomedicine & Pharmacotherapy, 2018, 103: 1652-1657.

[11] Huang Y, Li Y, Zhang K, et al. Expression and diagnostic value of PIWI-interacting RNA by serum in acute myocardial infarction[J]. Journal of Cardiology, 2023, 82(6): 441-447.

[12] Palazzo A F, Lee E S. Non-coding RNA: what is functional and what is junk?, J. Frontiers in genetics, 2015, 6: 2.

[13] Zhang K, Li Y, Huang Y, et al. PiRNA in cardiovascular disease: focus on cardiac remodeling and cardiac protection, J. Journal of Cardiovascular Translational Research, 2023, 16(4): 768-777.

[14] Vella S, Gallo A, Nigro A L, et al. PIWI-interacting RNA (piRNA) signatures in human cardiac progenitor cells, J. The international journal of biochemistry & cell biology, 2016, 76: 1-11.

[15] Lipps C, Northe P, Figueiredo R, et al. Non-invasive approach for evaluation of pulmonary hypertension using extracellular vesicle-associated small non-coding RNA, J. Biomolecules, 2019, 9(11): 666.

[16] Wu X, Han Y, Li J, et al. piRNA-5938 can regulate cardiomyocyte apoptosis and mitochondrial fission, J. Chinese Journal of Tissue Engineering Research, 2023, 27(11): 1750.

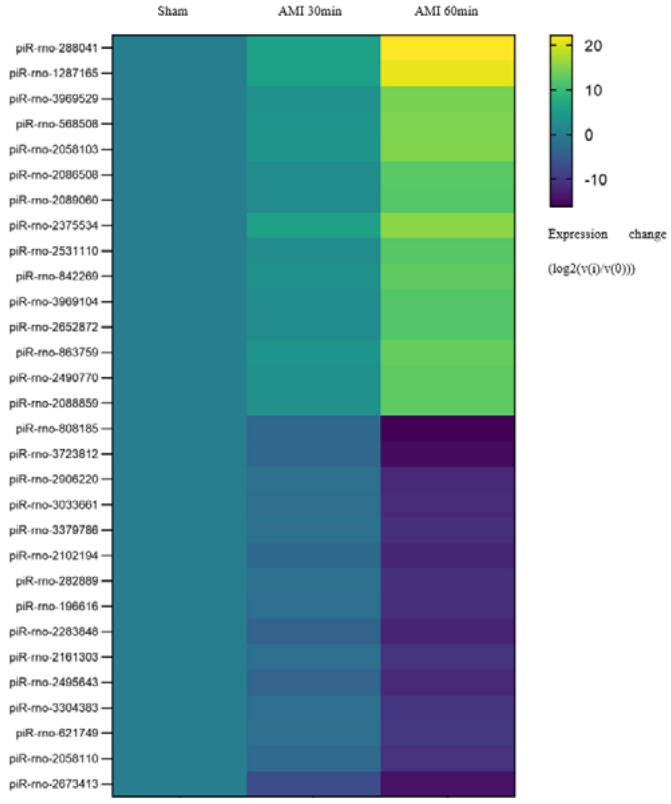


Figure 1. The AMI-related piRNAs in plasma exosomes from high-throughput sequencing. The x-axis shows the groups and the y-axis shows the category of the piRNAs.

Table 1. Target Genes Associated with piRNAs in AMI

Trends in piRNA changes with prolonged AMI onset/duration	PiRNA	Target Gene
Increase	piR-rno-288041	Nkain2, U4
	piR-rno-568508	Wwp2, Mir140
	piR-rno-2490770	Mterf3, 5S_rRNA
Decrease	piR-rno-3033661	Klrb1b
	piR-rno-282889	Mterf3, 5S_rRNA

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 929-935
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Advancing Forensic Age Estimation: Insights from Fluorescence Intensity and Time-Resolved Fluorescence Spectroscopy on Human Semen Stains

Sasithon Santikulthani[1], Natakorn Sapermsap[2], Sorawis Sangtawesin[1,2], Panomsak Meemon[1,2], Piyamas Petcharoen[1,3*]

[1]*Integrated Science and Innovation Program, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, Thailand.*

[2]*School of Physics, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, Thailand.*

[3]*School of Biology, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, Thailand.*

** Corresponding author at: School of Biology, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, Thailand, piyamas@sut.ac.th*

Abstract

Biological evidence, including blood, semen, and sweat stains, is crucial in forensic investigations by providing DNA for the identification of suspects or victims. Equally vital is the estimation of the age of these body fluid stains, offering critical insights into the timeline of a crime. This study employed two fluorescence spectroscopy techniques, namely fluorescence intensity and time-resolved fluorescence, to estimate the age of human semen stains. The semen stains deposited on glass surfaces for various durations (0, 1, 2, 3, 5, 7, and 14 days). Preliminary results indicated that fluorescence intensity exhibited significant changes in the long term, whereas time-resolved changes occurred in the short term. Trend line for these techniques were calculated using logistic and exponential models. The fluorescence intensity technique produced R-squared values at 0.86, indicating relatively strong predictive accuracy, but showing trends of possible convergence or divergence, suggesting the need for additional data to enhance accuracy in the future study. Conversely, the time-resolved fluorescence technique produced high R^2 values at 0.89 and convergence trending, indicating a more robust predictive capability. The integration of fluorescence lifetime with fluorescence intensity in the prediction model showed substantial potential for estimating the age of semen stains. Moreover, this methodology shows promise in estimating the age of other biological evidence and could potentially be incorporated into forensic light sources (FLS) as equipment for estimating body fluid age in the future.

Keywords
Age estimation, Human semen, Intrinsic fluorescence, Time-resolved fluorescence, Fluorescence lifetime

1. Introduction

The age of body fluids is crucial for determining event timelines, evidence integrity, and witness credibility [1]. Various techniques, including chemical [2]DNA [3], RNA [4], and protein analyses [5], are used to estimate body fluid age. Recently, in order to avoid contamination, non-contact and non-destructive techniques have been highly preferred [6]. The spectroscopic technique offers a fast, non-destructive method for identifying body fluids by analyzing their characteristic spectral signatures [7]. Fluorescence spectroscopy technique, for instance.

The fluorescence spectroscopy technique consisting of fluorescence intensity and time-resolved fluorescence was focused. The fluorescence intensity measures the emitted light as a function of wavelength, offering insights into the quantum yield and concentration of fluorescent molecules or fluorophores [8]. The time-resolved fluorescence spectroscopy provides information about fluorescence lifetimes based on detection of intensity decay, effectively used to elucidate structure and function of biological samples, e.g., proteins [9]. The difference between these two techniques is the nature or properties of the fluorophore and the consequent time of detection [10]. These techniques are powerful to study the intrinsic fluorescence property of human semen stains which is the biological evidence most found at sexual assault scene [11].

This work reports the preliminary results of utilizing fluorescence intensity and time-resolved fluorescence spectroscopy to study the intrinsic fluorescence properties of human semen, which varied with time.

2. Materials and methods

2.1. Sample collection and preparation

This project was conducted according to the Human Research Ethic Committee of Suranaree University of Technology (HREC-SUT, EC-66-0077). Human semen was collected from a healthy volunteer.

Human semen sample was collected by masturbation from ejaculation. The semen sample was pipetted with 15 l, deposited on a glass slide with three

replicates, and left to dry before measurement. During the waiting time condition (Days 1, 2, 3, …, 14) to measure, the samples were kept in ambient light at room temperature.

2.2. Sample measurement

The samples were performed with a custom-made fluorescence spectrometer including fluorescence intensity spectroscopy and time-resolved fluorescence spectroscopy. The custom-made fluorescence spectrometer was built within the Quantum Microscopy Laboratory (QLAB) at Suranaree University of Technology, Thailand. The fluorescence intensity spectroscopy consists of an excitation laser, optical fiber, bandpass filters and detector. The time-resolved fluorescence system consists of light source of a supercontinuum laser, dichroic filters, mirror, optical fiber, detector, single-photon counter and Time-correlated single photon counter (TCSPC). This measurement occurs in the dark room at room temperature. The parameter condition of fluorescence intensity spectroscopy consists of excitation at 405 nm with record emissions between 300 and 1000 nm. The parameter condition of time-resolved fluorescence spectroscopy consists of excitation at 405 nm with a signal delay of 10 ps.

2.3. Data analysis

Data analysis of fluorescence intensity technique and time-resolved fluorescence technique was conducted through maximum normalized intensity and fluorescence lifetime parameter versus time, respectively. Simple regression analysis was used.

3. Results

3.1. Changing in semen stain color on a glass slide

Human semen stain color changes from colorless on day 0 to increasingly yellow on days 2, 3, and continuing through day 14, with the intensity proportional to time.

3.2. Fluorescence intensity spectroscopy

The fluorescence intensity measurements of semen stain taken from 1 donor over a period of 14 days are presented in Figure 1a. The fluorescence intensity of semen was measured at an excitation wavelength of 405 nm, targeting fluorophores present in the semen stain. One potential fluorophore is the coenzyme protein

flavin adenine dinucleotide (FAD), which has an excitation range between 405 nm and 480 nm [12]. Maximum normalized intensity at emission wavelength of 460-500 nm was used and the results showed that fluorescence intensity increased over time, and rapid changes were observed between day 5 and day 14. The best fit was a logistic equation with an R-squared value of 0.86, which demonstrated relatively strong predictive accuracy. After day 14, showing trends of possible convergence or divergence. The predicted equation is below.

$$y = \frac{0.02 - 617.94}{1 + (x/649.27)^{2.28}} + 617.94$$

The predicted equation of *the fluorescence intensity* technique, where y is *maximum normalized intensity* and x is time in units of day.

3.3. Time-Resolved Fluorescence Spectroscopy

The time-resolved fluorescence measurements of semen stain taken from 1 donor over a period of 14 days are presented in Figure 1b. Excitation of 405 nm excites FAD in semen stain. Fitting raw data with exponential decay, fluorescence lifetime or fluorescence decay parameter was emphasized. The results indicate that fluorescence lifetime increased from 1.0 to 2.1 ns initially, converging around 2.1 ns, which rapid changes occur in the short term, particularly between day 0 and day 3. The best fit was an exponential association equation with an R-squared value of 0.89, which shows the predictive equation was strong. The predicted equation as

$$y = 1.09 + 0.46\left(1 - e^{-x/1.44}\right) + 0.56\left(1 - e^{-x/0.0008}\right)$$

which y is fluorescence lifetime in units of ns and x is time in units of day.

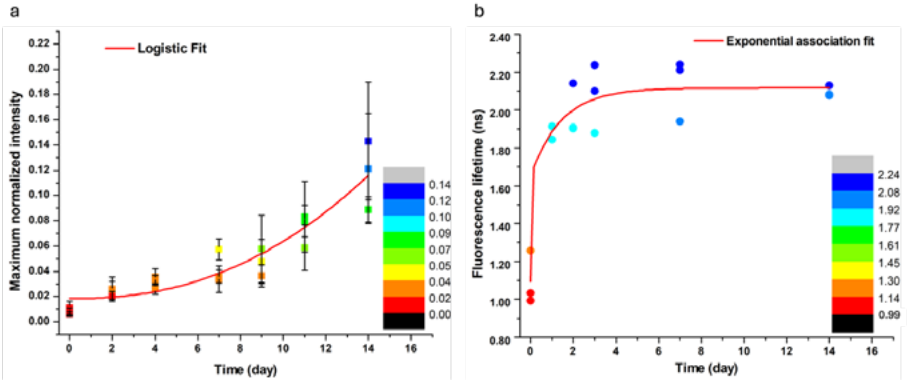


Figure 1a shows the correlative between the maximum normalized intensity with gradient color low to high (black to gray) and time of the human semen stain with the trend line of logistic modeling. **Figure 1b** shows the correlation between fluorescence lifetime parameter and time of human semen stain with trend line of exponential association modeling.

4. Discussion

In this work, using fluorescence intensity spectroscopy and time-resolved fluorescence spectroscopy to study intrinsic fluorescence properties of human semen stain on glass slide over time from day 0 to day 14. The fluorescence intensity of semen stain increases over time, which correlates with the color change of semen stain resulting from the redox and oxygenation reactions of nicotine adenine dinucleotide (NADH) and FAD [13]. Although the trend line gives a high R-squared value, the trending after day 14 can converge and diverge, not definitively established at present. The behavior of the logistic model is growth at first; once the population has grown to reach its environment's maximum capacity, it will level off around the carrying capacity [14]. In addition to the behavior of body fluids, it should remain stable over an extended period of time. To prove that and make it clearer, the experiment requires a longer time period, more than 14 days, at least 28 days, and more semen donors. The fluorescence lifetime of semen stains exhibits natural behavior; however, the results are based on a single donor. To refine the predictive equation, further data from multiple donors is required.

5. Conclusion

This project suggests that fluorescence intensity and time-resolved fluorescence spectroscopy are effective tools for distinguishing between fresh and old stains. However, further studies involving a larger donors and varying time intervals are required to accurately determine the exact time since deposition. Moreover, this procedure can be extended to analyze other body fluids, including blood and saliva. Utilizing predictive modeling in future research and development efforts to create a portable age estimation device.

6. Acknowledgments

Funding of this research was provided by the Development and Promotion of Science and Technology Talents Project (DPST).

7. Conflict of interest statement

The authors declare no conflicts of interest

8. References

[1] K. Alkass, B. A. Buchholz, S. Ohtani and et al, Age Estimation in Forensic Sciences, *Mol. Cell. Proteomics*, 9 (2010) 1022–1030.

[2] J. Qu, W. Meador, P. Cheah and et al, Latent bloodstain detection using a selective turn-on NIR fluorescence dye responsive to serum albumin, *RSC Adv.*, 13 (2023) 27549–27557.

[3] H. Nakanishi, A. Takada, K. Yoneyama and et al, Estimating bloodstain age in the short term based on DNA fragment length using nanopore sequencer, *Forensic Sci. Int.*, 358 (2024) 112010.

[4] J. Fu and R. W. Allen, A method to estimate the age of bloodstains using quantitative PCR, *Forensic Sci. Int. Genet.*, 39 (2019) 103–108.

[5] D. Bhuptani, S. Kumar, M. Vats, and et al, Age and gender related changes of salivary total protein levels for forensic application, *J. Forensic Odontostomatol.*, 36 (2018) 26–33.

[6] R. H. Bremmer, K. G. De Bruin, M. J. C. Van Gemert and et al, Forensic quest for age determination of bloodstains, *Forensic Sci. Int.,* 216 (2012) 1–11.

[7] F. Zapata and I. Gregorio, Body Fluids and Spectroscopic Techniques in Forensics: A Perfect Match?, *J. Forensic Med.,* 1 (2016).

[8] J. R. Lakowicz, Principles of fluorescence spectroscopy, Second., Boston, 2006.

[9] A. Boreham, R. Brodwolf, K. Walker and et al, Time-Resolved Fluorescence Spectroscopy and Fluorescence Lifetime Imaging Microscopy for Characterization of Dendritic Polymer Nanoparticles and Applications in Nanomedicine, *Molecules*, 22 (2016) 17.

[10] M. Y. Berezin and S. Achilefu, Fluorescence Lifetime Measurements and Biological Imaging, *Chem. Rev*., 110 (2010) 2641–2684.

[11] R. E. Gaensslen, Forensic Analysis of Biological Evidence, *Forensic Sci. Res*, 1 (2000).

[12] P. Vishnyakova et al., Fluorescence lifetime imaging microscopy as an instrument for human sperm assessment, *Biochem. Biophys. Res. Commun.,* 645 (2023) 10–16.

[13] S. Y. Lim, J. I. Jang, H. Yoon and et al, Spectroscopic Study of Time-Varying Optical Redox Ratio in NADH/FAD Solution, *J. Phys. Chem. B*, 126 (2022) 9840–9849.

[14] C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, An Introduction to Logistic Regression Analysis and Reporting, *J. Educ. Res.,* 96 (2002) 3–14.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 937-943
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Assessing the Utility of STR Profiling for Combating COVID-19 Insurance Fraud Involving. Lateral Flow Test Kits in Thailand

Korapin Srisiri[a], Panjit Ratanasathien[b], Achirapa Bandhaya[a*]

[a]*Forensic Science Unit, Faculty of Science, Mahidol University, Bangkok, Thailand.*
[b]*Mahidol University International College, Mahidol University, Nakhon Pathom, Thailand.*

** Corresponding author at: Forensic Science Unit, Faculty of Science, Mahidol University, Bangkok, Thailand. Email: achirapa.ban@mahidol.ac.th, Tel: +66 2 201 5212, Fax +66 2 354 7096*

Abstract

During the COVID-19 pandemic in Thailand, insurance companies introduced policies that paid benefits to policyholders who tested positive for the virus. At the peak of the outbreak, accessing quantitative real-time PCR tests and results was challenging, causing delays in treatment and insurance claims. Consequently, the government allowed self-tested positive results from Lateral Flow Immunochromatographic Assays (LFIAs) to be used as proof for issuing medical certificates. This created a vulnerability for insurance fraud, as individuals could use another person's positive LFIA result to obtain a certificate to claim benefits.

This study explored the feasibility of authenticating saliva-based LFIAs for SARS-CoV-2 antigen using the conventional workflow of Short Tandem Repeat (STR) profiling, with a focus on the impact of time since sample deposition on DNA viability. The yield of DNA recovered from three LFIA areas were assessed, and the quality of profiles generated from samples that had been stored for 1, 7, 14, 30 days, and six samples aged between 6 to 17 months were evaluated for the number of alleles obtained and the heterozygous peak height balance. The results showed that complete STR profiles could be generated from all 30 LFIAs in this study, indicating that adequate amount of high-quality DNA was still recoverable even after over one year of sample deposition. These data suggested that STR profiling could be used as a tool for law enforcement and insurance companies to combat fraudulent claims associated with LFIAs for COVID-19 testing.

Keywords

Lateral Flow Immunochromatographic Assays, DNA profiling, STR, COVID-19, Insurance fraud.

1. Introduction

The COVID-19 pandemic posed unprecedented challenges to global healthcare systems and economies. In Thailand, the rapid spread of the virus required swift adaptations in medical testing and insurance claim protocols. As quantitative real-time PCR testing resources became strained due to high demand, the Thai government authorised the use of self-administered SARS-CoV-2 antigen lateral flow immunochromatographic assays (LFIAs), known locally as Antigen Test Kits (ATKs), for official medical certification. This enabled insurance claims to be filed based on positive LFIA results. However, the policy inadvertently facilitated fraudulent practices, allowing individuals to use another person's positive LFIA to claim insurance benefits. The financial damage from this and other types of COVID-19 insurance fraud exceeded 500 million Baht [1].

LFIAs are rapid, point-of-care diagnostic tools designed to detect specific antigens or antibodies, commonly used in infectious disease testing [2]. A LFIA typically includes a test strip with several key components: a sample pad for applying the specimen, a conjugate pad containing labelled monoclonal antibodies that bind to the target antigen, a nitrocellulose membrane where test and control lines form, and an absorbent pad to capture excess fluid. In SARS-CoV-2 antigen tests, the kit generally includes a plastic cassette housing the test strip and a buffer solution for sample preparation. Collection devices, such as swabs or containers, along with a pipette or dropper for sample transfer, are often provided, depending on whether the kit is designed for nasopharyngeal, nasal, or saliva samples.

Previous works have successfully performed Short Tandem Repeat (STR) typing on DNA recovered from used swabs and buffer solutions of LFIAs intended for body fluid identification, such as blood and saliva [3]. However, no attempts have yet been reported to extract and profile DNA from test strips of used LFIAs for the purpose of human identification, particularly those designed for disease detection.

This study aimed to assess the feasibility of using STR profiling to verify the donor of the specimen deposited on SARS-CoV-2 antigen LFIAs. To develop this protocol, DNA was extracted from various areas of the used LFIA strips, and the section yielding the highest DNA quantity was selected for further analysis. STR profiling was performed on LFIAs stored for different durations, including samples aged between 6 and 17 months, to examine DNA persistence on this substrate. The profiles were evaluated based on the number of recoverable alleles, allelic dropout, and heterozygous peak height balance. The findings provide proof-of-concept for

the use of STR profiling to authenticate samples on COVID-19 LFIAs and potentially on LFIAs for other antigens, offering a tool to mitigate disputes and reduce insurance fraud.

2. Material and methods

The study protocol was approved by Mahidol University Central Institutional Review Board under the Certificate of Approval number MU-CIRB 2024/224.0309.

2.1. Sample Collection and Preparation

To identify LFIA segments that retain the highest amount of DNA, a healthy volunteer self-administered a COVID-19 test using Flowflex Antigen Rapid Test kit (ACON Biotech), designed to detect SARS-CoV-2 antigens in saliva, following the manufacturer's instructions. After air-drying at room temperature for one day, the LFIA cassette was opened, and approximately 1 cm from the sample pad, nitrocellulose membrane, and absorbent pad areas of the test strip were collected into separate 1.5-mL centrifuge tubes for DNA extraction. Three test replicates were performed on three separate days.

To assess the impact of time since sample deposition on DNA stability and viability, three healthy volunteers were instructed to collect approximately 3 mL of saliva into an envelope supplied with the LFIA kit. Following the manufacturer's protocol, they aliquoted the saliva into two LFIA extraction buffer tubes, each to the specified level (approximately 1 mL), using the provided plastic pipette. After thorough mixing by squeezing and inverting the tubes, four drops of the saliva-buffer mixture were applied to the sample area of the LFIA cassette. Each volunteer submitted four sets of two LFIAs, with each set collected on different days, to the researchers. These LFIAs sets were stored for 1, 7, 14, or 30 days at room temperature (24-29°C) and relative humidity (61-76%). After each storage period, approximately 1 cm of the LFIA section containing the highest amount of DNA was collected into 1.5-mL centrifuge tubes for extraction. Additionally, six used LFIAs stored at room temperature for 6, 10, 12, 13, 15, and 17 months, donated by one of the volunteers, were processed as described above.

Buccal swabs were also collected from each volunteer to generate reference profiles.

2.2. DNA Extraction and quantification

DNA extraction from different sections of the LFIA strips and buccal swabs was performed using a modified version of the QIAamp DNA Micro Kit (QIAGEN)'s protocol, as described by Srisiri et al. [4]. DNA was eluted with 20 µL of Buffer AE, and the quantity was measured using the Qubit 3.0 Fluorometer and Qubit dsDNA HS Assay Kits (Thermo Fisher Scientific).

2.3. STR Amplification and Fragment Analysis

Approximately 0.5-1 ng of extracted DNA was used as the template for STR amplification using AmpFℓSTR Identifiler Plus PCR Amplification Kit (Thermo Fisher Scientific). The amplified products were analysed on an Applied Biosystems 3500 Genetic Analyzer and data were processed with GeneMapper ID-X Software v1.6 (Thermo Fisher Scientific). The analytical threshold was set at 125 RFU. Profiles from each volunteer's LFIAs were compared with their respective reference genotypes for the number of alleles detected. The heterozygous peak height balance was also evaluated.

3. Results and discussion

DNA quantification revealed that the sample pad section of the LFIA consistently yielded the highest amount of recoverable DNA across all samples. In contrast, negligible amount of DNA was obtained from the nitrocellulose membrane and absorbent pad areas. This finding aligns with the expectation that the sample pad, being the initial site of sample deposition, is likely to retain the majority of biological material.

The scatter plot in Figure 1 illustrates the relationship between DNA concentration and average allele peak heights across different storage periods for LFIAs processed from three volunteers. An inconsistent correlation is observed between DNA yield and average peak height; some samples with high DNA concentrations exhibited lower peak heights, while others with low concentrations produced higher peak heights. For example, one replicate of sample from Volunteer 1, stored for 14 days, showed similar DNA concentrations but lower peak heights compared to the second replicate from the same storage period. Additionally, some samples from Volunteer 2 with DNA concentrations as low as ~0.2 ng/µL generated profiles with average peak heights exceeding 1000 RFU.

The time of day at which samples were collected also appeared to affect DNA yield. Samples collected in the morning, prior to drinking or toothbrushing, showed elevated DNA concentrations. This may be due to the co-extraction of oral microbial genetic material along with epithelial DNA, which could have led to an overestimation of the total DNA yield, potentially masking the actual amount of human DNA present [5]. To address this issue, future analyses could employ quantitative real-time PCR (qPCR) to distinguish human DNA from microbial or other non-human DNA.

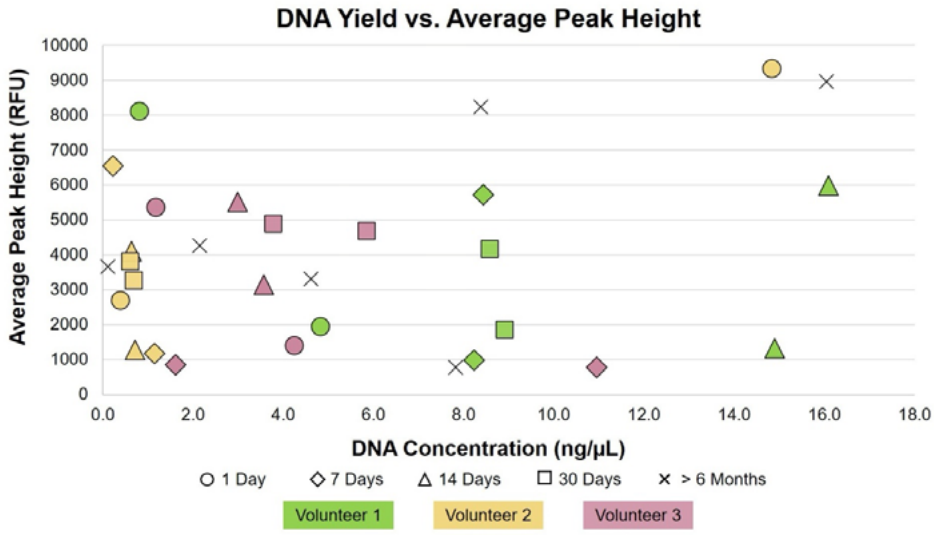


Figure 1. A scatter plot showing the DNA quantity and average peak heights for LFIA samples collected from three volunteers across different storage periods. Each data point represents one replicate. All six LFIA samples aged over six months were donated by Volunteer 2.

Despite the modest correlation between the DNA yield and the average peak height of the corresponding sample profiles, full STR profiles were successfully generated from all LFIAs in the study, including those stored for 6 to 17 months at room temperature. The allele peak heights in these profiles were well-balanced, with average peak height ratios ranging from 72% to 92% (Figure 2). This suggests that even under ambient indoor conditions, DNA remains sufficiently intact for successful amplification.

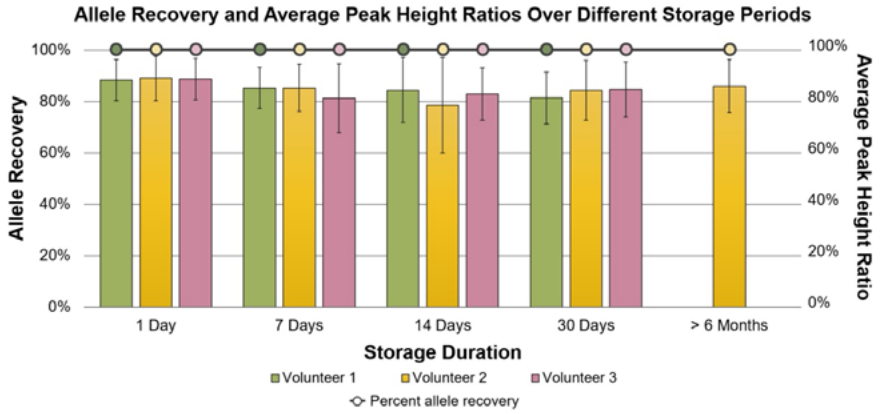


Figure 2. A chart illustrating the allele recovery and average peak height ratios from LFIA samples stored over different time periods, analysed across three volunteers. Each bar represents the mean value from two replicates, with error bars indicating the standard deviation. Percent allele recovery (circles) is displayed on the primary y-axis, while average peak height ratios (bars) are shown on the secondary y-axis.

4. Conclusions

This proof-of-concept study demonstrates the feasibility of using STR profiling to authenticate the origin of specimens deposited on used SARS-CoV-2 antigen LFIAs. Complete STR profiles were successfully generated from LFIAs stored for up to 17 months, highlighting the stability of DNA on this substrate over extended periods. No consistent correlation was observed between DNA quantity and the average peak heights. Quantifying DNA using qPCR should help improve the estimation of the amount of human DNA available in each sample, however, this would also cause a rise in the cost of analysis. With the increasing use of LFIAs for disease and drug detection, their portability and ease of use make them valuable tools in various contexts. The successful application of STR profiling on these test strips could help address issues such as insurance fraud, where verifying the integrity and identity of samples is critical.

5. Acknowledgments

The presentation of this research project at the 30th Congress of the ISFG was partially supported by the Faculty of Science and the Faculty of Graduate Studies, Mahidol University. The funding source had no role in any aspect of this study or this paper.

6. Conflict of interest statement

None.

7. References

[1] False Covid-19 insurance claims in excess of THB500 million, [Internet] The Nation Thailand, 2022 Jul 5 [cited 2024 Aug 25], Available: https://www.nationthailand.com/in-focus/40017388.

[2] G. Li, Q. Li, X. Wang, et al. Lateral flow immunoassays for antigens, antibodies and haptens detection, Int. J. Biol. Macromol. 242 (2023) 125186.

[3] S.C. Zapico, G. Roca, Making the most of lateral flow immunochromatographic tests: an efficient protocol to recover DNA, Methods Protoc. 7 (2024) 8.

[4] G. Sguazzi, H.L. Mickleburgh, S. Ghignone, et al. Microbial DNA in human DNA extracts: recoverability of the microbiome in DNA extracts stored frozen long-term and its potential and ethical implications for forensic investigation, Forensic Sci. Int. Genet. 59 (2022) 102686.

[5] J. Rosenbaum, M. Usyk, Z. Chen, et al. Evaluation of Oral Cavity DNA Extraction Methods on Bacterial and Fungal Microbiota. Sci. Rep. 9 (2019) 1531.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 945-953
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Assessing the impact of inhibitors on the viability of biological samples for forensic DNA analysis: A pilot study

Adeyanju Mosuro[a], Sharlize Pedroza-Matute[a,b], Steve Cummings[b], Tasnim Munshi[a], Sasitaran Iyavoo[a,b*]

[a] *University of Lincoln, Lincoln, United Kingdom.*
[b] *AttoLife Limited, Norwich, United Kingdom.*

** Corresponding author at: School of Natural Sciences, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, United Kingdom, siyavoo@lincoln.ac.uk.*

Abstract

The value of DNA evidence in forensic cases is substantial; however, the presence of inhibitors in biological samples can pose challenges for forensic scientists. This study aimed to assess the impact of inhibitors on forensic DNA analysis, focusing on the limitations of current DNA quantification and amplification protocols. K562 Genomic DNA was diluted to eight concentrations (100, 33, 11, 3.7, 1.23, 0.41, 0.14, and 0.05 ng/µL) and combined with two common inhibitors: haematin (in four concentrations: 2, 20, 200, and 2000 µM) and humic acid (in three concentrations: 15, 150, and 1500 ng/µL). DNA quantification was performed using the Quantifiler™ Trio DNA Quantification Kit, and amplification was carried out with the VeriFiler™ Express PCR Amplification Kit. Inhibition was assessed by the failure or delay of Internal Positive Control amplification, and the Degradation Index was calculated. DNA profiling effects were evaluated based on the quality of electropherograms. The results indicated that at their highest concentrations, both haematin and humic acid inhibited DNA quantification, causing either complete amplification failure or delayed IPC amplification beyond 30 threshold cycles. Samples with 1000 µM haematin showed DI values greater than 1, while 750 ng/µL of humic acid caused amplification failure of small and/or large DNA targets. The inhibitory effect increased in samples containing 16.5 and 50 ng/µL DNA combined with haematin, yet the quality of electropherograms remained unaffected. Similarly, humic acid exhibited minimal inhibitory effects, though a reduction in peak height for larger base pairs was noted at its highest concentration. In conclusion, this study highlights the significant impact of inhibitors on DNA quantification over amplification and underscores the potential inhibitory effects of high DNA concentrations.

Keywords
Forensic DNA analysis, PCR inhibitors, Quantifiler™ Trio kit, haematin, humic acid, DNA quantification.

1. Introduction

Since the first use of DNA fingerprinting for forensic DNA analysis in 1986, the field of forensic genetics has moved towards more straightforward and efficient methods. To date, the most popular DNA profiling method is Short Tandem Repeat (STR) typing. STRs are DNA loci characterised by short repetitive units of DNA that form alleles with varying lengths. These alleles are most commonly typed via Polymerase Chain Reaction (PCR), a technique that enables the targeted amplification of DNA fragments, followed by separation by length through capillary electrophoresis. Using this method, DNA found in biological fluids at crime scenes can be typed and compared to the DNA of suspects, providing invaluable support to criminal investigations [1].

Despite the popularity of PCR-based STR typing, this method has certain limitations. While the technique works well in the presence of pure, high-quality, and sufficient quantities of DNA, obtaining good samples is not always straightforward, given the nature of biological stains found at crime scenes. PCR is sensitive to various types of contaminants that may be present in biological matrices and the environment. These substances, known as PCR inhibitors, can interfere with different PCR components and processes, leading to suboptimal or failed amplification, even when a good quantity of non-degraded DNA is available. Common forensic PCR inhibitors include haematin found in blood, humic acid found in soil, collagen found in bones, melanin found in hair, dyes found in fabric, and certain carryover substances from DNA extraction [2-3].

In forensic genetics, PCR is not only used for STR typing but is also frequently employed for DNA quantification. Determining the amount of DNA available for STR typing is often fundamental to ensuring effective STR profiling, and one of the most effective methods for this is real-time quantitative PCR [4]. The utility of this technique in genetic testing also extends to evaluating the presence of inhibitors, allowing for the determination of whether further purification steps may be necessary before proceeding with STR typing. Inhibition can decrease the efficiency of amplification at each PCR cycle, reducing the slope of the curve in the exponential phase and delaying the cycle at which the signal surpasses the threshold

[5]. Some commercial kits, such as the Quantifiler™ Trio DNA Quantification Kit, are designed to detect inhibition through the addition of an internal positive control (IPC). The IPC is a synthetic DNA template expected to cross the threshold with a consistent threshold cycle (Ct) in non-inhibited samples. In the presence of inhibition, the IPC Ct is expected to increase. The Quantifiler™ Trio kit is also designed to detect a small autosomal (SA) and a large autosomal (LA) target in the human genome, allowing for the identification of degradation. This is because, in degraded DNA, the larger target is selectively depleted compared to the shorter target. The SA/LA ratio, referred to as the degradation index (DI), increases in more degraded samples [6].

This study aims to evaluate the effect of two common inhibitors, haematin and humic acid, on DNA quantification using the Quantifiler™ Trio kit, and amplification using the VeriFiler™ Express PCR Amplification Kit, a STR kit with high discrimination power and the capability of direct amplification [7]. Different levels of DNA input in combination with the contaminants are also tested to assess whether the effect of inhibitors is influenced by the availability of DNA.

2. Material studied, methods, techniques

2.1. Study samples

K562 Genomic DNA (Promega, supplied at a concentration of 400 µg/mL in TE buffer) was diluted to eight different concentrations (100, 33, 11, 3.7, 1.23, 0.41, 0.14, and 0.05 ng/µL) with deionised water. Stock solutions of haematin (Sigma-Aldrich®) and humic acid (Sigma-Aldrich®) were prepared using 1 N sodium hydroxide and deionised water [8]. Then, four concentrations of haematin (2, 20, 200, 2000 µM) and three concentrations of humic acid (15, 150, 1500 ng/µL) were separately combined with the DNA dilutions in equal volumes. Therefore, the final concentration of DNA and inhibitors in the tested samples was half of the initial concentration reported.

2.2. DNA quantification and amplification

Each sample was quantified using the Quantifiler™ Trio DNA Quantification Kit *(Applied Biosystems™)*, following the manufacturer's recommendations. The same samples were amplified using the VeriFiler™ Express PCR Amplification Kit *(Applied Biosystems™)*, adhering to a reduced volume protocol [7]. All samples were processed in triplicates.

2.3. Data analysis

The effect of inhibition was evaluated by reviewing the Quantifiler™ Trio IPC Ct values. The DI was also calculated as the ratio between the concentrations of the SA and LA targets. The results were compared with the data reported in the Quantifiler™ Trio validation study [6]. The inhibitory effects on STR typing with the VeriFiler™ Express kit were determined based on the quality of the obtained electropherograms.

3. Results

As shown in Table 1 and 2, both haematin and humic acid, at their highest concentrations, either prevented amplification or delayed the IPC Ct values beyond 30. Higher IPC Ct values were also observed in association with high DNA concentrations: 16.5 and 50 ng/µL of DNA combined with any amount of haematin resulted in IPC values close to or exceeding 30, similar to the results obtained when combining 50 ng/µL of DNA with any concentration of humic acid.

A DI higher than 1 was observed in samples containing 1000 µM of haematin. For samples containing 750 ng/µL of humic acid, the DI was undetermined due to the amplification failure of the SA and/or LA targets.

The results from DNA quantification did not correspond to those observed in STR typing. The quality of the electropherograms remained unaffected by haematin at any concentration, irrespective of the DNA quantity. In contrast, limited inhibitory effects were seen when typing samples containing humic acid, although a reduction in peak heights for larger base pairs was noted at the highest concentration.

Table 1. Average degradation index (DI) and internal positive control (IPC) threshold cycle (Ct) values obtained from Quantifiler™ Trio quantification of the triplicate study samples, determined by a combination of DNA and haematin at the indicated final sample concentrations. Where one of the triplicate samples had an undetermined DI, the average DI value is indicated with a '>' sign before the highest DI value, as the undetermined value would be even higher.

DNA concentration (ng/µL)	Haematin concentration (µM)	DI	IPC Ct
0.025	0	0.68	29.64
	1	0.59	29.74
	10	0.57	29.64
	100	0.49	29.95
	1000	>21	30.60
0.07	0	0.47	29.69
	1	0.52	29.69
	10	0.55	29.70
	100	0.59	29.84
	1000	>9	30.91
0.205	0	0.53	29.49
	1	0.58	29.44
	10	0.58	29.64
	100	0.52	29.68
	1000	4.88	31.16
0.615	0	0.52	29.49
	1	0.48	29.44
	10	0.47	29.32
	100	0.50	29.55
	1000	58.65	35.07
1.85	0	0.57	29.41
	1	0.53	29.30
	10	0.57	29.38
	100	0.57	29.68
	1000	1.70	34.50

5.5	0	0.55	29.53
	1	0.53	29.53
	10	0.56	29.75
	100	0.52	29.78
	1000	3.17	Undetermined
16.5	0	0.59	29.90
	1	0.59	29.96
	10	0.57	30.12
	100	0.54	30.20
	1000	1.88	Undetermined
50	0	0.68	30.34
	1	0.64	30.25
	10	0.64	30.43
	100	0.57	30.57
	1000	2.78	Undetermined

Table 2. Average degradation index (DI) and internal positive control (IPC) threshold cycle (Ct) values obtained from Quantifiler™ Trio quantification of the triplicate study samples, determined by a combination of DNA and humic acid at the indicated final sample concentrations.

DNA concentration (ng/µL)	Humic Acid concentration (ng/µL)	DI	IPC Ct
0.025	0	0.26	29.11
	7.5	0.41	28.93
	75	0.51	28.59
	750	Undetermined	Undetermined
0.07	0	0.50	28.58
	7.5	0.55	28.53
	75	0.51	28.57
	750	Undetermined	Undetermined
0.205	0	0.51	28.44
	7.5	0.75	28.03
	75	0.39	28.32
	750	Undetermined	Undetermined

0.615	0	0.52	28.29
	7.5	0.65	28.12
	75	0.56	28.11
	750	Undetermined	Undetermined
1.85	0	0.66	28.03
	7.5	0.71	27.89
	75	0.56	28.29
	750	Undetermined	Undetermined
5.5	0	0.79	28.40
	7.5	0.71	28.48
	75	0.53	28.70
	750	Undetermined	Undetermined
16.5	0	0.71	28.90
	7.5	0.70	28.91
	75	0.57	29.29
	750	Undetermined	Undetermined
50	0	0.79	29.27
	7.5	0.68	29.45
	75	0.59	30.07
	750	Undetermined	Undetermined

4. Discussion

The results obtained from the real-time PCR quantification were comparable to the data reported in the Quantifiler™ Trio validation study [6]. In that study, haematin concentrations of 200 to 1250 µM were used, showing delayed IPC values and amplification failure of SA and/or LA targets starting from 750 µM. In the present study, a gap was present between 100 to 1000 µM of haematin, with similar challenges observed at 1000 µM and acceptable results at 100 µM. Likewise, in the validation study, humic acid concentrations of 200 to 800 ng/µL were used, with issues starting from 400 ng/µL. The present study showed comparable results when using 750 ng/µL, and minimal effects with the second highest concentration of 75 ng/µL. Another study produced similar results using different real-time PCR-based quantification kits [9].

This study, therefore, confirmed previous observations, including high DI values being linked to inhibition rather than DNA degradation, as good quality

control DNA was used. Additionally, this investigation revealed increased inhibition associated with higher DNA amounts in the sample. In fact, excessive amounts of DNA are considered a PCR inhibitor, leading to poor DNA amplification, primarily due to physical obstructions [10]. It is perhaps not surprising to observe increased inhibition in the samples analysed in this study, where a combination of high inhibitor concentrations and high DNA amounts was present.

Although inhibitory effects were observed during DNA quantification, this study reported very little effect on the STR typing results. This suggests an enhanced resilience to inhibitors in the VeriFiler™ Express kit compared to the Quantifiler™ Trio kit. A separate study also demonstrated limited inhibitory effects of haematin (250 to 750 µM) and humic acid (100 to 300 ng/µL) when using four other forensic STR kits [11].

Future investigations may be required to fully understand the effects of inhibitors on PCR for both DNA quantification and amplification, including increasing the number of samples, adding more inhibitor data points, and testing a wider range of inhibitor types.

5. Conclusion

This study demonstrated the effects of different concentrations of haematin and humic acid on real-time PCR quantification using Quantifiler™ Trio kit and STR typing using the VeriFiler™ Express kit, highlighting a greater impact on forensic DNA quantification than on amplification. The inhibitory effect of high DNA concentrations in combination with elevated levels of contaminants was also emphasised. However, further studies are required to expand our understanding of inhibition and to work towards improving quantification and amplification methods.

6. Conflict of interest statement

All claims expressed in this article are solely those of the authors and do not necessarily represent the views of their affiliated company or organisation.

7. References

[1] U. Nwawuba Stanley, A. Mohammed Khadija, A.T. Bukola, et al., Forensic DNA Profiling: Autosomal Short Tandem Repeat as a Prominent Marker in Crime Investigation. Malays. J. Med. Sci. 27 (2020) 22–35.

[2] K. Vajpayee, H.R. Dash, P.B. Parekh, et al., PCR inhibitors and facilitators – Their role in forensic DNA analysis. Forensic Sci. Int. 349 (2023) 111773.

[3] R. Alaeddini, Forensic implications of PCR inhibition – A review. Forensic Sci. Int. Genet. 6 (2012) 297-305.

[4] J.A. Nicklas, E. Buel, Quantification of DNA in forensic samples, Anal. Bioanal. Chem. 376 (2003) 1160-1167.

[5] E.J. Kontanis, F.A. Reed, Evaluation of Real-Time PCR Amplification Efficiencies to Detect PCR Inhibitors, J. Forensic Sci. 51 (2006) 795-804.

[6] A. Holt, S.C. Wootton, J.J. Mulero, et al., Developmental validation of the Quantifiler® HP and Trio Kits for human DNA quantification in forensic samples, Forensic Sci. Int. Genet. 21 (2016) 145-157.

[7] J. Perry, T. Munshi, T. Haizel, et al., Validation of reduced volume VeriFiler™ Express PCR Amplification Kit for buccal swab samples extracted using Prep-n-Go™ Buffer. J. Forensic Sci. 67 (2022) 1971-1978.

[8] Q. Hu, Y. Liu, S. Yi, et al., The effect of six common PCR inhibitors on DNA polymerase and DNA template, Int. J. Forensic Sci. Pathol. 2(8) (2014) 1-4.

[9] A.S. Holmes, R. Houston, K. Elwick, et al., Evaluation of four commercial quantitative real-time PCR kits with inhibited and degraded samples, Int. J. Leg. Med. 132 (2017) 691-701.

[10] M.L. Altshuler, PCR Troubleshooting: The Essential Guide, Caister Academic Press, Norwich, 2006.

[11] W. Cho, J.K. Jung, D.G. Lee, et al., A comparative evaluation of forensic STR kits and their tolerance to PCR inhibitors. Aust. J. Forensic Sci. (2023) 1-10.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 955-961
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Implementation of an AI system for automated sperm identification

Albin Mobäck[1], Linda Erlandsson[1], Malin Rapp[1], Therése Gunnarsson[1], Ricky Ansell[1,2], Christina Forsberg[1,3], Arvid H. Gynnå[1*]

[1] *National Forensic Centre, Swedish Police Authority, Linköping, Sweden*
[2] *Department of Physics, Chemistry & Biology, Linköping University, Sweden*
[3] *Applied Microbiology, Department of Chemistry, Lund University, Sweden*

** Corresponding author at: arvid.heden-gynna@polisen.se, +46 10-56 24 077*

Abstract

Microscopy for identification of spermatozoa is a labor-intensive task in forensic laboratories. Automation of imaging and image interpretation has the potential to greatly improve throughput. Automation may also increase quality by helping to ensure consistency between operators or laboratories.

We describe the evaluation, validation and ongoing implementation of MetaSystems Metafer, an automated microscope system which includes robotic feeding of microscopic slides and automated image analysis by artificial intelligence. The system can be loaded with up to 80 microscopic slides, read the slides unsupervised, assign a score to potential sperm in each slide and present those to an operator in order ranked by likeness to a sperm cell. The operator can then make a judgement on whether to accept the presented objects as sperm or not.

We have investigated how to optimize specificity and sensitivity with regards to the operator effort needed. It was found that imaging in multiple focus planes considerably increases specificity for spermatozoa. We have also mapped the distribution of scores for sperm and non-sperm objects, allowing the laboratory to decide on a well-informed policy on how many potential sperm objects that must be reviewed to find positive samples with sufficient sensitivity. Whether certain sample types are unsuitable for automated analysis is also investigated. Finally, we have performed an in-house validation, finding that sensitivity using the automated system surpasses that of manual microscopy.

Keywords

Sperm detection, microscopy, artificial intelligence, automation, sexual assault.

1. Introduction

At the Swedish National Forensic Center, the examination of 6 000 microscopic slides each year for sperm is a labor-intensive task. To increase sample throughput, we wish to introduce automated imaging and image interpretation. Automation may also contribute to improved quality by reducing variability between operators.

For this purpose, we used a robotic microscope with AI-based automated sperm recognition from MetaSystems Hard & Software GmbH (Altlussheim, Germany). We have investigated how to optimize specificity and sensitivity with regards to the operator effort needed, and performed an internal validation of the system. The use of the system in our laboratory has been accredited and we are currently finalizing implementation.

2. Material studied, methods, techniques

2.1. Microscope

For microscopy, an AxioImager Z2 upright microscope with motorized stage and objectives EC Plan-Neofluar 20x/0.5 and EC Plan-Neofluar 40x/0.75 were used (Carl Zeiss AG, Oberkochen, Germany). Images were acquired with a CoolCube 4c camera (MetaSystems). The system was also equipped with a slide feeding robot, a magazine for 80 microscopic slides and a bar code reader (MetaSystems). Microscope control and image analysis used a computer equipped with the Metafer software version 4.3.5 (MetaSystems).

Sample wells (see 2.2) were imaged in a patchwork pattern using the 20x objective. Objects were detected using the method *FSPD-OD-V1.10* and scored using the method *FSPD-OC-V2.11-Madrid,* both available in the Metafer software. Objects are scored for similarity to a sperm cell on a scale ranging from zero to one hundred points. When multiple z-planes were acquired, a merged image was created using the *PyramidFusion* method.

2.3. Sample preparation

The 188 microscopic slides used were acquired from routine case examinations and prepared according to standard procedure at the laboratory. Briefly, a subsample of a swab or cutting is placed on a microscope slide in a 14 mm (Assistent, Karl Hecht GmbH) or 11 mm (Marienfeld Superior) diameter well. The sample is dissected manually with needles in the well, and then dried. Alternatively, a sub-

sample may be taken during differential extraction after removal of the epithelial fraction and placed in a 6 mm well (Marienfeld Superior). The sample is stained by Oppitz' Christmas tree method by adding Nuclear fast red (Kernechtrot) for 15 minutes, rinsing in water, adding Picroindigocarmine for 15 seconds and rinsing with 95 % ethanol [1]. Slides are then dried and a cover slip is mounted using glucose syrup mounting media.

3. Results

3.1. Microscopy optimization

Initial experiments were performed to optimize the performance of the system. While sperm cells were generally found to receive high scores by the algorithm, there were also many non-sperm objects being scored highly. We found that imaging in multiple z-planes, and performing object detection and scoring on the merged image, increased the specificity of the scoring algorithm for sperm. Specifically, imaging three or four planes (at +2, 0, -2 µm or +1, -1, -3 and -5 µm relative to the focus plane) decreased the number of objects scoring 80 points or higher by more than two thirds, while preserving all actual sperm (Figure 1A). Imaging in two planes (0, -3 µm), however, was found to decrease sensitivity.

To determine how sperm were scored in the different imaging modes, six samples with 75 sperm cells in total were imaged in a single plane, three planes or four planes (Figure 1B). The cumulative score distribution indicated that the four-plane mode yielded higher scores for sperm cells, and it was thus selected for further use.

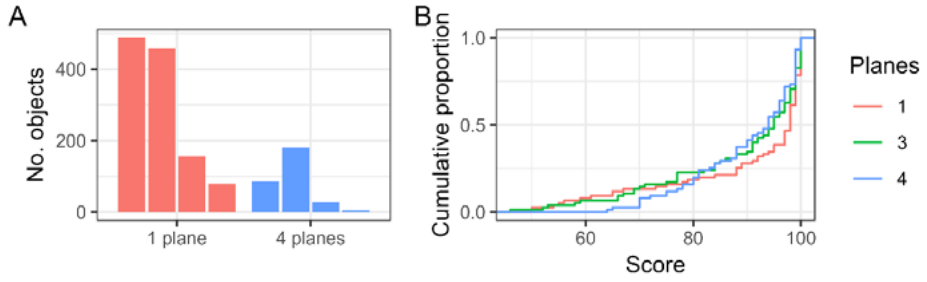


Figure 1. A. Number of objects scoring at least 80 points in four samples depending on imaging mode. B. Cumulative distribution of scores for 75 sperm cells in six samples, depending on imaging mode.

3.2. Distribution of scores

While the majority of sperm cells were found to score above 90 points, others with scores as low as 52 points were also observed. Additionally, even though stacked imaging improved specificity, non-sperm object with high scores were still frequently observed. Generally, it cannot be assumed that an actual sperm present in a sample will be the highest scoring object. To determine how many objects that need to be reviewed by an operator to achieve an acceptable rate of false negatives, we decided to map how the scores assigned to sperm and non-sperm were distributed.

To determine the distribution of scores for sperm, 34 samples with 1-15 sperm cells as determined by manual microscopy (in total 152) were imaged. With the automated system, 248 sperm were found with scores ranging from 52 to 100 points (Figure 2A). The 10th percentile was at 77 points, and 87.5% were at 80 points or above.

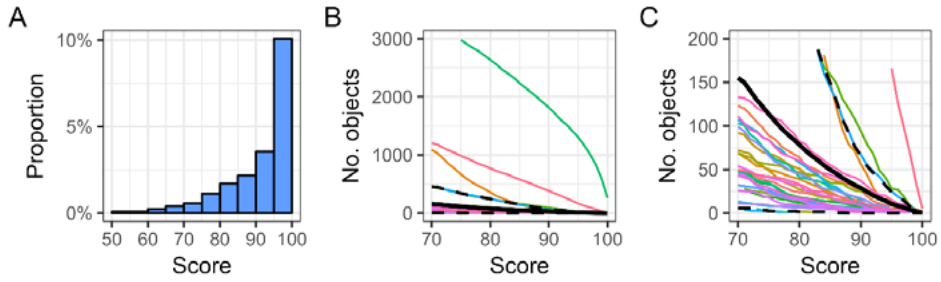


Figure 2. A. Distribution of scores for 248 sperm cells in 34 samples. B. Cumulative number of non-sperm objects at or above each score value in 32 samples. Each colored line represents one sample. The solid black line is the average with max and min values excluded, and the dashed lines the 10th and 90th percentiles. C. As B, showing up to 200 objects only.

For non-sperm objects, 32 slides deemed negative by manual microscopy were imaged and the number of objects at or above a certain score were mapped (Figure 2B-C). The number of objects per slide varied greatly, and while most had 30 or less objects at or above 80 points, a few samples contained hundreds or thousands of such objects.

Aiming to optimize both sensitivity and specificity, we deemed that reviewing only objects scoring in the 80-100 interval would be likely to provide both sufficient sensitivity for positive samples and an acceptable operator workload. As this

interval included 87.5% of the sperm cells, samples containing two sperm cells are expected to be found positive with 98.4% sensitivity.

3.3. Validation of the method

As part of the method validation, we investigated sensitivity, performance on challenging samples, repeatability, inter-operator variability and selectivity when reviewing proposed sperm on-screen.

To determine sensitivity for presence of sperm cells, 169 samples, of which 84 had previously been found positive by manual microscopy, were examined using the automated system. The positive samples contained various levels of sperm and epithelial cells, and 21 samples were specially selected for being challenging to examine, e.g. due to high levels of debris, focusing problems, condom lubricants or animal sperm. With the automated system, a single sperm was found in four samples and five sperm cells in one sample previously found negative by the manual method. The latter sample had been deemed difficult to evaluate, and differential extraction had been ordered despite a negative microscopy result. These five brought the number of positive samples to 89. In total, the sensitivity of the manual method was found to be 94.4%, while the automated system found 96.6% of all samples considered positive by either method. Four samples were considered unsuitable for examination with the automated system, either because of focus problems, stemming mainly from bubbles in the cover slip mounting media, or an abundance of lubricant droplets from condoms.

To evaluate repeatability, five slides were imaged twice and the score of ten objects in each was recorded. For sperm cells, the mean deviation was one point and the maximum observed seven points. Non-sperm objects saw larger differences, on average 26 points and maximum 39 points. For inter-operator variability, it was found that two operators agreed completely on sperm presence or absence when examining 27 samples independently.

Suggested sperm are shown on-screen where the operator may scroll through the image stack to view different z planes. The microscope stage can also navigate directly to the position of the object, allowing it to be viewed through the ocular. As it is quicker to confirm sperm cells directly on screen, we wanted to evaluate whether on-screen review was as specific as viewing through the ocular. To this end, 150 objects from 17 samples considered to be sperm cells on-screen were double-checked through the microscope ocular. All 150 objects were determined to be sperm also through the ocular.

4. Discussion

Artificial intelligence has previously been shown to be useful for sperm detection [2]. Here, we report an automated system passing all validation parameters and thus being considered fit for the purpose of sperm detection in a forensic laboratory. Certain types of samples were considered unsuitable for the system: samples with focusing problems, samples with high amounts of debris and condom samples. However, all samples can be imaged automatically initially. If sperm are found, they can be reported as positive. If negative and considered unsuitable, they can either be subjected to manual microscopy, or differential extraction can be begun and a new slide made after removal of the epithelial fraction.

The process of using the automated system includes three steps: batch setup, imaging and review of proposed sperm objects. The setup time is mostly independent of the number of samples included. The imaging time depends on the imaged area and the number of z planes; four planes in a 14 mm well takes about 13 minutes. Importantly, the automated imaging can be performed without user intervention, e.g. overnight. The time needed for review is determined by the number of objects the operator has to review, and whether possible sperm can be confirmed directly on-screen. We found that by mapping the distributions of scores on observed sperm cells, a score-wise threshold can be set that limits which objects that need to be reviewed while retaining sufficient sensitivity. A threshold of 80 points was found to fulfill these requirements in our laboratory. We have also concluded that sperm may be confirmed on screen if the operator is confident in the identification, and we expect a majority of samples to be handled such. The time needed for review is then estimated to 1-4 minutes per sample.

Aside from increased throughput, automated microscopy is also expected to increase quality in sperm detection. While it is shown here that sensitivity is higher with the automated system, we also expect inter-operator consistency to be improved by partial automation. The recording of sperm will also make operator training and synchronization easier, as examples and good practices can now more easily be shared.

While the system was found suitable and efficient for the detection of sperm, its implementation also includes challenges. The imaging and reviewing of samples are most efficiently done in a batch manner, while slides previously have been examined per-case. The adaption of a workflow that allows us to optimally benefit from the system is an ongoing effort.

5. Conclusion

The system was found to be fit for purpose, passed accreditation, and will be implemented in routine practice. Careful evaluation of imaging conditions and the scoring algorithm allowed us to achieve a high sensitivity together with a minimal operator hands-on time per sample. A remaining challenge is to adapt our laboratory workflow to the batch-wise operation of the instrument, where samples are preferentially imaged at night and reviewed during the next day.

6. Conflict of interest statement

None.

7. References

[1] E. Oppitz, *Eine neue Färbemethode zum Nachwies der Spermien bei Sittlichkeitsdelikten*, Ark. Kriminologie, 144 (1969) 145–148

[2] R. Golomingi, C. Haas, A. Dobay et al., *Sperm hunting on optical microscope slides for forensic analysis with deep convolutional networks – a feasibility study*, Forensic Science International: Genetics, 56 (2022) 102602

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 963-969
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Automation and LIMS Development for the PowerSeq™ CRM Nested Assay

Michelle Galusha[1], Michael Brandhagen[1], Jodi Irwin[1], Robyn Wolfe[1*]

Federal Bureau of Investigation Laboratory, 2501 Investigation Parkway, Quantico, VA, USA 22135

** Corresponding author: rpwolfe@fbi.gov*

Abstract
The following studies were performed by the FBI Laboratory to internally validate the automation scripts and laboratory information management system (LIMS) integration of the Promega PowerSeq® CRM Nested System, which was previously validated for manual laboratory processing. Validation of the automated workflow included development, testing and validation of the robotic scripts used for laboratory processing together with integration of the entire system in the LIMS, STACs. Overall, these studies serve to establish the reliability of the automated PowerSeq™ mtDNA CR Next Generation Sequencing (NGS) System for mtDNA CR typing.

Keywords
Mitochondrial DNA; automation; next generation sequencing

1. Introduction

Though automated processes and laboratory information management systems (LIMS) are routinely employed for CE STR typing, few reports describe automation of mtDNA processes and/or NGS processes. For mtDNA, very few labs have the type of case volume requiring automation and, for those labs that do, automation has been primarily restricted to the sequencing of high quality and quantity reference samples that can typically be handled in a more standardized fashion than low quantity/quality evidence samples [1](Taylor et al. 2020). For NGS STRs, automated processes for both high- and low- quality samples have been described [2, 3] (Montano et al. 2018), Hollard et al. 2019); but because STR testing is less prone

to contamination, STR workflows tend to be more amenable to automation than mtDNA workflows. Here, we report on the development of automated protocols and LIMS modules for next generation mitochondrial DNA sequencing of evidence specimens using the previously validated PowerSeq CRM Nested Assay [4](Brandhagen et al. 2018).

2. Materials and Methods

2.1. Automation/LIMS testing and optimization

Automated protocols were developed on Tecan workstations, with the pre-PCR steps of mtDNA qPCR and PowerSeq CRM amplification performed on a Tecan Evo, and the post-PCR steps of bead purification, PowerSeq Quant MS set-up and DNA library normalization and pooling performed on a Tecan Fluent. Tecan scripts were developed based on Promega's published protocol and the FBI's manual validation of the assay (Brandhagen et al. 2018).

The automated workflow was integrated with the laboratory information system in use by the FBI, STACS 5.1. Custom STACS modules were designed for each step of the PowerSeq CRM process (e.g., extraction, extract quantification, indexed CRM amplification, amplicon purification, pooling, etc.). Additionally, the LIMS provides sample/batch tracking via unique barcodes, reagent tracking, file import/export functionality with instrumentation, and data import/storage of mtDNA profiles. In addition, a specific reagent tracking utility was developed to monitor individual index combinations and index volumes to ensure that each index combination is unique to its respective reaction.

Calculations required at various steps in the workflow were computed by STACS. Calculations included: averaging of extract replicate quantities, computation of degradation index values, normalization of amplicons, determination of stock amplicon quantity from multiple diluted replicates, stock amplicon normalization for pooling, and reagent volume calculation for sequencing.

2.2. Workflow Validation Experiments

2.2.1. Accuracy and Precision

Accuracy and precision studies used 24 samples (23 buccal swabs and one DNA control) with known mtDNA CR profiles previously developed by Sanger and/or NGS sequencing, and which represented 24 distinct CR haplotypes. All samples were collected and typed with informed consent.

2.2.2. Sensitivity

Sensitivity studies were performed with a dilution series of a sample with a known CR sequence at the following ten DNA inputs: 20,000, 10,000, 5,000, 2,500, 1,250, 625, 312, 156, 78, 39 mtDNA copies.

2.2.3. Case-type samples

Five non-probative hairs and eleven non-probative calcified tissue samples along with 31 associated reagent blanks were processed. Multiple RBs were processed with each sample to develop data to assess background signal and contamination with the hair and bone laboratory workflows employed in routine operational casework.

2.2.4. Contamination

Possible sources of contamination were assessed using known samples, reagent blanks (RBs), positive controls and negative controls (NCs). The experiments were designed to evaluate both environmental contamination and contamination introduced by the automated script. Environmental contamination was assessed by typing RBs and NCs, while contamination introduced by the automated process was evaluated by alternating RBs and/or NCs with known samples to assess aerosol/well-to-well contamination.

3. Results

Accuracy and Precision

Across the three experiments performed, the 675 expected variants expected among the 24 samples were observed across the 79,400 bp analyzed.

To assess the precision of the automated workflow, variant frequency percentages for the 675 variants were compared in pairwise fashion across the three experiments. For substitutions and point heteroplasmies, differences in variant frequency percentages averaged 0.61% and were <0.7% in 75% of comparisons. For the more variable length heteroplasmies, differences in frequency percentages averaged only 1.19% (Figure 1).

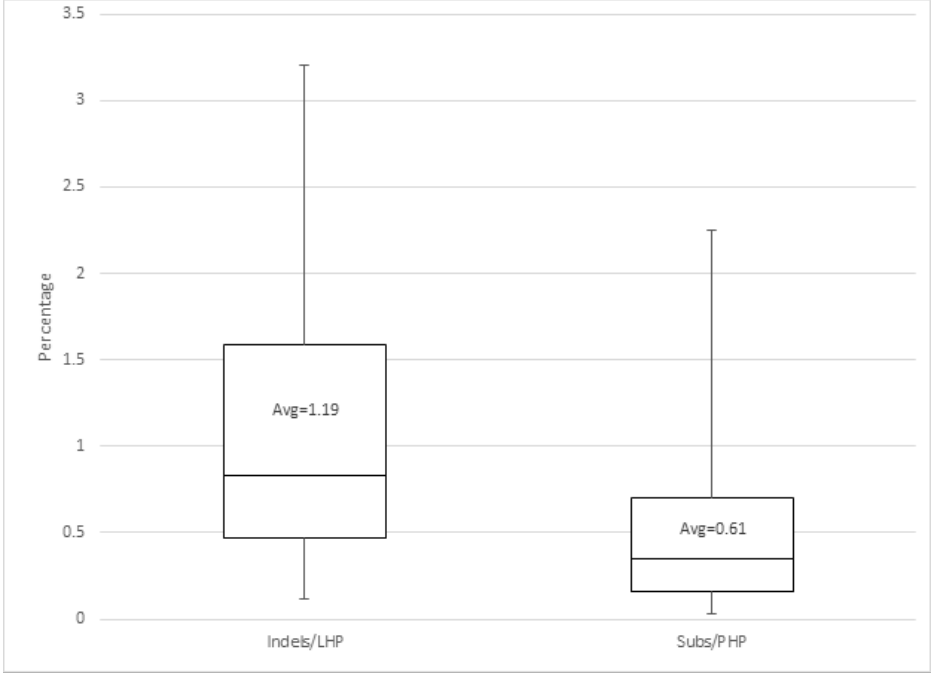


Figure 1. Percentage difference in variant frequency percentages observed across the three runs of 24 samples.

Sensitivity

Complete control region profiles were recovered down to 156 copies of mtDNA input. At 78 copies, 85% of the CR was recovered, while at 39 copies, 56% of the CR was recovered.

Case-type samples

Three hair samples, amplified with between 1100 and 6200 copies and all with low degradation indices produced full CR profiles. Two hair samples, amplified with 0 and 50 copies of input according to qPCR [5] and both with undetermined degradation indices, yielded no data. Extreme degradation likely explains the lack of data with 50 copies from hair "Franc" (Table 1).

Table 1. Sample characteristics and sequencing metrics for the five hair samples tested.

Sample Description	Total mtDNA Copies Used	Degradation Index	% recovered	min (reads)	max (reads)	avg (reads)
Franc	51.5	undet	0	0	78	18
49	6,259	0.47	100	2,296	51,719	16,809
38	3,501	0.59	100	1,873	48,858	16,195
1060	1,171	0.93	100	129	7,595	2,235
DSU0189	0	undet	0	0	6	0

All calcified tissue samples and amplification replicates produced full CR profiles with inputs ranging from 3,000-25,000 copies. However, for one sample, inconsistent and mixed data were observed in both amplification replicates. As the sample exhibited a high degradation index (>5), the sequencing results likely reflect DNA damage.

Contamination

Environmental contamination was assessed with 50 negative controls and reagent blanks processed in the absence of DNA-containing samples. For these experiments, average read depth was 8 reads, indicating a minimal level of background contamination.

When DNA containing samples were introduced to the experiments, the average read depth of the negative controls and reagent blanks increased substantially. Experiments were thus performed to minimize cross-contamination caused by the presence of high quantity DNA. Dilutions of high quantity/quality samples prior to loading them on the deck successfully reduced read counts in the NCs and RBs. When contamination was re-assessed with ~200 controls processed alongside with 1:50 diluted samples, the NC/RB data supported the use of a 200 read depth analytical threshold.

4. Discussion

A comprehensive automated workflow was developed for mtDNA processing of evidentiary specimens, with pre-PCR steps performed on a Tecan Evo and post-PCR steps performed on a Tecan Fluent. The robotic protocols were fully integrated with

the FBI LIMS, STACS®. LIMS modules were designed and customized to provide, among other things, sample and batch tracking via unique barcodes, reagent tracking, import/export functionality with instrumentation, and data import/storage of mtDNA profiles. Optimization of the system resulted in a final automated workflow that included the addition of longer bleach holds for the fixed tip robot, pipette mixing of purification beads and dilution of high quantity samples prior to loading on the robot.

5. Conclusion

Automated protocols and LIMS modules for the PowerSeq CRM assay were successfully developed, tested, and optimized. Validation of the final automated workflow established the accuracy, reproducibility, sensitivity, and reliability of the LIMS integrated system - from DNA extraction through final NGS haplotype review and LIMS storage.

6. Acknowledgments

The authors would like to thank Anthony Onorato, Susannah Kehl, Tavis Sipe, Spencer Rohrbacher, Danielle Daniels, and Odile Loreille for their support for and assistance with this work.

7. Disclaimer

Names of commercial manufacturers are provided for identification purposes only, and inclusion does not imply endorsement of the manufacturer, or its products or services by the FBI. The views expressed are those of the authors and do not necessarily reflect the official policy or position of the FBI or the U.S. Government.

8. References

[1] C.R. Taylor, K.M. Kiesler, K. Sturk-Andreaggi, J.D. Ring, W. Parson, M. Schanfield, P.M. Vallone, C. Marshall, Platinum-quality mitogenome haplotypes from United States population, Genes 11 (2020) 1290.

[2] E.A. Montano, Bush, A.M. Garver, M.M. Larijani, S.M. Wiechman, C.H. Baker, M.R. Wilson, R.A. Guerrieri, E.A. Benzinger, D.N. Gehres, M.L. Dickens, Optimization of the Promega PowerSeq™ Auto/Y system for efficient integration within a forensic DNA laboratory. Forensic Sci Int Genet. 32 (2018) 26-32.

[3] C. Hollard, L. Ausset, Y. Chantrel, S. Jullien, M. Clot, M. Faivre, É. Suzanne, L. Pène, F.X. Laurent, Automation and developmental validation of the ForenSeq™ DNA Signature Preparation kit for high-throughput analysis in forensic laboratories. Forensic Sci Int Genet. 40 (2019) 37-45.

[4] M.D. Brandhagen, R.S. Just, J.A. Irwin, Validation of NGS for mitochondrial DNA casework at the FBI Laboratory. Forensic Sci Int Genet. 44 (2020)

[5] M.F. Kavlick, Development of a triplex mtDNA qPCR assay to assess quantification, degradation, inhibition, and amplification target copy numbers. Mitochondrion 46:41-50.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 971-977
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Enhanced Interpretation of Complex DNA Mixtures Using Multiple Highly Polymorphic MPS-based Microhaplotype Panels

Zhuo Li[1], Mengyu Tan[2], Jiaming Xue[2], Xiao Zhang[1], Shengqiu Qu[2], Bo Jin[1*], Weibo Liang [2*]

[1]*Institute of Basic Medicine and Forensic Medicine, North Sichuan Medical College, Nanchong, 637000, Sichuan, China*

[2]*Department of Forensic Genetics, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu, 610041, Sichuan, China*

**Corresponding authors Fax number: +86 28 85401825*

E-mail address: liangweibo@scu.edu.cn, jinbo@nsmc.edu.cn.

Abstract

The interpretation of DNA mixtures remains a significant challenge in forensic science, particularly when dealing with complex mixtures involving multiple contributors, low DNA template quantities, or highly imbalanced ratios. In this study, we employed massively parallel sequencing (MPS) to assess the performance of three highly polymorphic microhaplotype panels in resolving complex DNA mixtures. The panels included a 55-plex (average effective allele number, A_e, of 7.43), a 67-plex (A_e of 5.34), and a combined 87-plex (A_e of 7.02).

A total of 56 artificial mixtures, consisting of DNA from 2 to 4 contributors in varying ratios and amounts, were analyzed. Detection rates were measured across different contributor proportions, DNA input levels, and numbers of contributors. Our results demonstrated that the three panels provided accurate and sensitive detection, even for low-template and highly imbalanced mixtures. The study highlighted the robustness of microhaplotype panels, particularly Panel A with its high polymorphism, in analyzing mixtures with multiple contributors, while Panel C performed exceptionally well in reducing allelic dropout and detecting minor contributors in imbalanced mixtures.

In summary, this study advances forensic practice by improving the effectiveness of DNA mixture interpretation through the application of MPS-based microhaplotype panels.

Keywords

Microhaplotype, High polymorphism, DNA mixture, Massively parallel sequencing.

1. Introduction

DNA mixtures are a common type of biological sample encountered at crime scenes. These mixtures contain the DNA of two or more individuals and are frequently associated with cases such as sexual assault, physical altercations, and homicides involving multiple participants. The highly informative nature of DNA mixtures makes their analysis crucial, as it can provide valuable insights regarding suspects, narrow the scope of investigation, and aid in the reconstruction of crime scenes. Consequently, numerous studies have focused on DNA mixture analysis[1].

However, the interpretation of DNA mixtures remains a significant challenge in forensic genetics due to uncertainties regarding the number and relative proportions of contributors. Traditional methods often struggle to accurately resolve mixtures, particularly when contributors are present in unequal amounts. To address this issue, researchers have explored various genetic markers to enhance the precision of mixture analysis.

In 2013, Kidd et al. introduced a novel genetic marker known as the microhaplotype. This marker is defined as a DNA fragment that contains two or more single nucleotide polymorphisms (SNPs) within a 300 bp region[2]. Microhaplotypes have since attracted considerable attention in forensic genetics, as they combine the advantages of short tandem repeats (STRs) and SNPs, offering benefits such as a large number of loci, broad distribution across the genome, high polymorphism, low mutation rates, and reduced recombination rates. Additionally, microhaplotypes are free from stutter artifacts during amplification, which is a common issue with STR analysis. These advantages render microhaplotypes particularly suitable for DNA mixture deconvolution[3].

Massively parallel sequencing (MPS) has emerged as an ideal technology for microhaplotype genotyping, as it allows for the phasing of each variant within a locus, enabling the discrimination of parental haplotypes. Furthermore, MPS is highly efficient, providing the throughput necessary to analyze complex mixtures with multiple contributors and imbalanced ratios[4].

2. Material and methods

2.1. Selection of candidate panels

We conducted a genome-wide screening to identify novel microhaplotype (MH) markers, consisting of two or more variants (either InDel or SNP) located

within a 300 bp region, based on data from the Southern Han Chinese population in the 1000 Genomes Project. The average effective allele number (A_e) was calculated using theoretical haplotype frequencies. Three highly polymorphic microhaplotype panels were selected for further analysis. These panels included a 55-plex panel (Panel A, with an average A_e of 7.43)[5], a 67-plex panel (Panel B, with an average A_e of 5.34)[6], and a combined 87-plex panel (Panels C, with an average A_e of 7.02) formed by merging the aforementioned two panels.

2.2. Sample preparation

A total of 56 artificial DNA mixtures, consisting of contributions from 2 to 4 individuals in various proportions, were prepared (Table 1). For each mixture ratio, four DNA input gradients were used: 1, 0.5, 0.25, 0.125, and 0.05 ng. Imbalanced mixtures were created by adding a large proportion of major contributor DNA to a fixed amount of minor contributor DNA at predetermined ratios.

Table 1. List of the number of contributors and the ratio of artificial mixture samples

Samples	NOC	Ratio	Contributors
Mix1-1	2	1:1	X-17/J-11
Mix1-2		1:5	
Mix1-3		1:10	
Mix1-4		1:20	
Mix1-5		1:40	
Mix2-1	3	1:1:1	X-17/J-11/W-36
Mix2-2		1:1:5	
Mix2-3		1:1:10	
Mix2-4		1:10:20	
Mix2-5		1:10:40	
Mix3-1	4	1:1:1:1	X-17/J-11/W-36/L-3
Mix3-2		1:1:3:5	
Mix3-3		1:3:5:10	
Mix3-4		1:5:10:20	

NOC: The number of contributors of artificial mixtures; Mix: Mixture

2.3. Sequencing and data analysis

The multiplex primer systems for candidate MHs in these panels were designed by iGeneTech Biotech Co. Ltd (Beijing, China). The DNA libraries were prepared using the MultipSeq® Custom Panel v3.0 (iGeneTech), following the

manufacturer's protocol for adaptor addition. The libraries were sequenced on the Illumina® NovaSeq 6000 platform using the 150 bp paired-end sequencing (PE150) mode. The FASTQ files generated from sequencing were analyzed using Trimmomatic v.0.38 and FastQC for quality control. Burrows Wheeler Aligner (BWA) v.0.7.12 and SAMtools were used to align the reads to the reference human genome (Hg19, GRCh37) and obtain binary alignment/map (BAM) files. An in-house pipeline for MH genotype calling was constructed using the CIGAR and MD: Z tag information of the BAM files.

To evaluate the performance of these panels in mixture analysis, the number and proportion of detected alleles were analyzed and compared with the expected alleles in each individual sample and mixture for all artificial mixtures. The genotypes of all contributors, which were known, were considered as the expected alleles, while the observed alleles in the mixture profiles were considered the observed alleles.

3. Results

The expected allele counts for the three panels are 174, 193, and 264 in a two-person mixture; 239, 247, and 352 in a three-person mixture; and 284, 295, and 430 in a four-person mixture. As the DNA input decreased and mixture imbalance increased, the number of observed alleles correspondingly reduced.

For two-person mixtures at a 1:40 ratio with an input amount of 1 ng, the detection rates for the minor contributor's alleles by the three panels were 95.98% (Figure 1A, Panel A), 96.89% (Figure 1B, Panel B), and 96.59% (Figure 1C, Panel C). In the case of three-person mixtures at a 1:10:40 ratio, the minor contributor's alleles were detected at rates of 95.82%, 91.90%, and 92.61%, respectively, across the three panels. For four-person mixtures at a 1:5:10:20 ratio, the detection rates were 99.30%, 97.29%, and 98.60%, respectively.

The sensitivity of allele detection was robust across all three panels, with detection efficiency reaching 0.5 ng in mixtures with equal ratios, even for four-person mixtures. At 0.5 ng, the panels could fully detect alleles in two-person mixtures at a 1:10 ratio, three-person mixtures at a 1:1:10 ratio, and a four-person mixture only at 1:1:1:1.

Even at an input amount as low as 0.05 ng, for two-person mixtures, the detection rates of the alleles of the minor donor by the three panels were 67.81% (Panel A), 72.54% (Panel B), and 70.45% (Panel C) at the 1:40 ratio, respectively.

For three-person mixtures, at the 1:10:40 ratio, the detection rates of the alleles of the minor donor by the three panels were 78.24%, 85.42%, and 80.97%, respectively. For four-person mixtures, at the 1:5:10:20 ratio, the detection rates of the alleles of the minor donor by the three panels were 82.39%, 87.46%, and 86.05%, respectively.

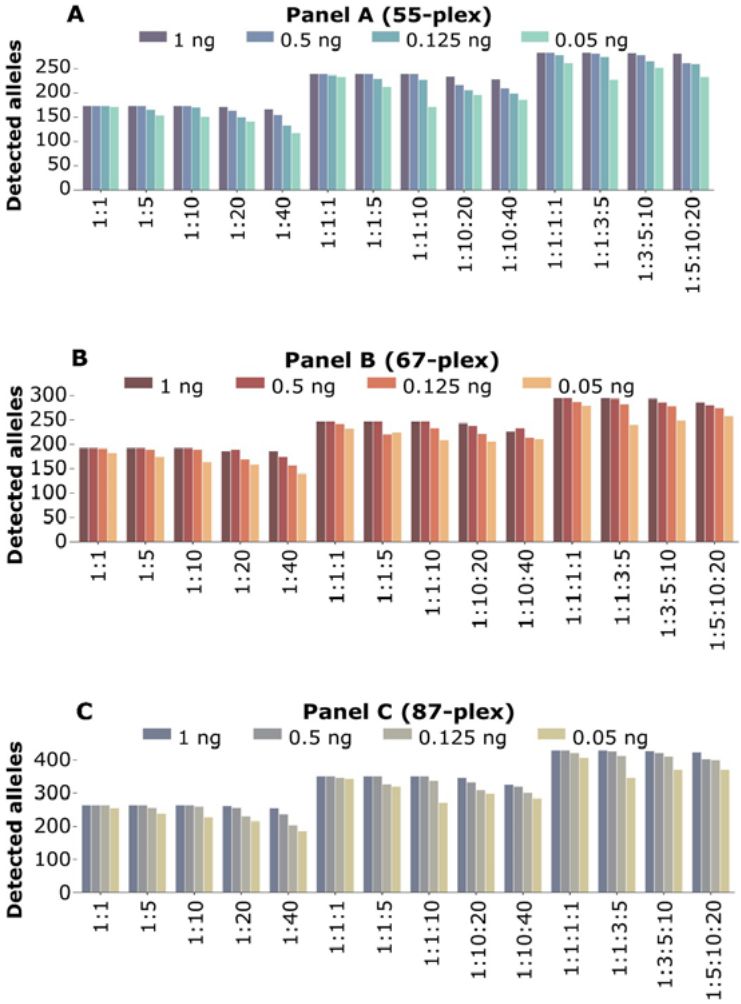


Figure 1. Allele detection in complex DNA mixtures using multiple panels (A: Panel A, B: Panel B, C: Panel C).

4. Discussion

In this study, we conducted a comparative analysis of mixtures with varying numbers of contributors, different mixture ratios, and different input template amounts using three panels with different polymorphic microhaplotypes. The results indicate that sequencing analysis of microhaplotype panels using MPS has great potential in the deconvolution of complex mixtures.

The three panels we tested showed considerable accuracy in analyzing mixtures with low DNA input and unbalanced contributor ratios. Specifically, Panel A, which exhibited the highest polymorphism, was particularly effective in analyzing mixtures with low-template DNA from multiple contributors. On the other hand, Panel C, with its higher number of loci, showed superior performance in reducing allelic dropout rates and providing more comprehensive information about minor contributors in highly imbalanced mixtures.

Our findings further suggest that the detection sensitivity of these panels is robust even at low DNA input levels. All three panels maintained high detection rates for minor contributors, even at input amounts as low as 0.05 ng. This highlights the potential of these panels in forensic investigations where DNA quantities are often limited or where contributors are present in vastly unequal proportions.

5. Conslusion

In conclusion, this study underscores the advantages of using MPS-based microhaplotype panels in forensic genetics, particularly for the interpretation of complex DNA mixtures. The application of these highly polymorphic panels enhances the resolution of DNA mixture deconvolution, offering improved accuracy in cases involving low DNA quantities and unbalanced contributor ratios. Future research should continue to explore the optimization of microhaplotype markers and sequencing technologies to further refine forensic DNA analysis.

6. Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 82271929), and Sichuan Science and Technology Program (No. 2023NSFSC1451).

7. Conflict of interest statement

None.

8. References

[1] Yang J, Lin D, Deng C, Li Z, Pu Y, Yu Y, *et al.* The advances in DNA mixture interpretation. *Forensic Science International 2019*, 301: 101-106.

[2] Kidd KK, Pakstis AJ, Speed WC, Lagace R, Chang J, Wootton S, *et al.* Microhaplotype loci are a powerful new type of forensic marker. *Forensic Science International: Genetics Supplement Series 2013*, 4(1): e123-e124.

[3] Oldoni F, Kidd KK, Podini D. Microhaplotypes in forensic genetics. *Forensic science international Genetics 2019*, 38: 54-69.

[4] Oldoni F, Bader D, Fantinato C, Wootton SC, Lagace R, Kidd KK, *et al.* A sequence-based 74plex microhaplotype assay for analysis of forensic DNA mixtures. *Forensic science international Genetics 2020*, 49: 102367.

[5] Tan M, Xue J, Wu Q, Zheng Y, Liu G, Zhang R, *et al.* Improving DNA mixtures analysis using compound markers composed of InDels and SNPs screened from the whole genome with next-generation sequencing. *Electrophoresis 2023*.

[6] Xue J, Tan M, Zhang R, Chen D, Liu G, Zheng Y, *et al.* Evaluation of microhaplotype panels for complex kinship analysis using massively parallel sequencing. *Forensic science international Genetics 2023*, 65: 102887.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 979-984
DOI: https://dx.doi.org/10.15304/cc.2025.1869

A Software Package for Designing and Interpreting Forensic DNA Validation Studies

Sarah Riman[1*], Hari Iyer[1], Edgar Robitaille[2], Sicen Liu[2], Asmitha Sathya[2], Ashwatha Suresh[2], Peter M Vallone[1]

[1]*National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD, USA.*
[2]*Johns Hopkins University, 3400 N. Charles St, Baltimore, MD, USA.*

** Corresponding author at: Applied Genetics Group, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA, Email: sarah.riman@nist.gov*

Abstract
Internal validation of measuring and interpreting single-source and mixture DNA profiles is essential in every forensic DNA laboratory. Validation is not a one-time process as laboratories continue to revalidate when changes or upgrades are introduced to their workflows. The field lacks much-needed open-source software that can assist in designing validation experiments and interpreting the resulting data.

Keywords
Internal validation, mixture profiles, open-source software.

1. Introduction

The internal validation process often generates massive amounts of data; forensic practitioners face the daunting challenges of designing validation experiments, formatting, analyzing, and understanding the validation data. Currently, software programs analyze (i.e., call alleles) [1, 2], deconvolute, and assign Likelihood Ratios (LRs) of DNA profiles [3, 4]. However, the field lacks a much-needed open-source tool that can assist in designing and interpreting the validation experiments. To address this critical need, we propose the development of a standalone (i.e., not running over a network) and easy-to-use graphical user interface (GUI) that can be easily integrated into the validation workflows and assist the practitioners in

automating (1) the design of validation samples to adequately cover the factor space, and (2) the interpretation and visualization of the data from validation studies.

2. Methods

The source codes for the different functionalities were developed in Python and R programming language formats that will be publicly available upon the release of the software. PROVEDIt data [5] and in-house datasets available in the NIST Applied Genetics Group are being used to test the codes.

3. Results

The GUI will consist of the following functionalities:

3.1. Construction of all possible mixture genotype combinations

The tool will support loading user-provided ground truth single-source profiles genotyped by any Short Tandem Repeat (STR) multiplex kit in a .txt file format. The tool will generate all possible mixture genotype combinations [6] depending on the number of single-source profiles and number of contributors (NoCs) chosen using the following formula:

$$Combinations, nCr = \frac{n!}{r!(n-r)!}$$

where ${}_nC_r$ = *number of possible genotype combinations*
n = number of single-source profiles
r = number of contributors

For example, to experimentally design mixture samples with NoCs = 2, 3, 4, and 5 from 11 different single-source DNA profiles, there will be 55, 165, 330, and 462 different mixture combinations, respectively.

3.2. Complexity metrics

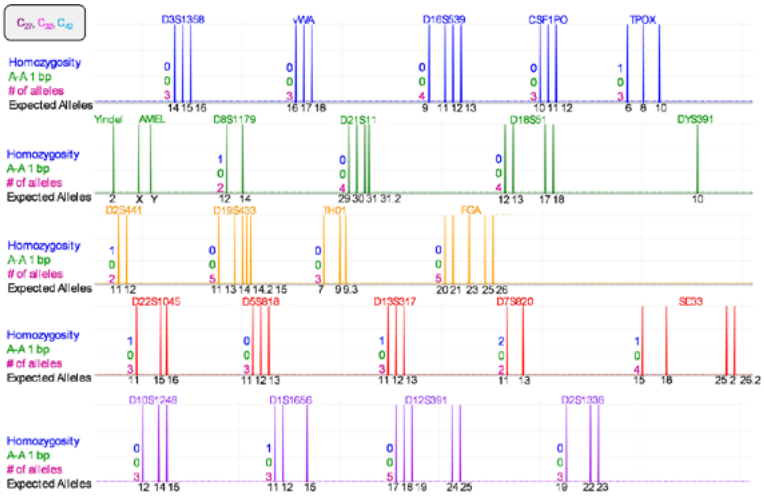
The tool will provide summary statistics for each of the different generated mixtures to aid the user with their choice of combinations (e.g., Fig. 1). These metrics will cover:

— expected alleles to be observed at each locus
— counts of alleles expected to be observed at each locus and across the entire profile

— minimum/maximum number of expected alleles for a profile
— instances of a single base-pair difference between two alleles at a given locus (referred to here as A-A (1 bp))
— counts of homozygote genotypes at a locus
— quantification of allele-sharing between the contributors that constitute each combination using allele-sharing ratio (ASR). The ratio will range between 0 (no sharing) to 1 (maximum sharing) and is generated using the following formula:

$$ASR = 1 - \frac{(actual\ number\ of\ observed\ peaks\ -\ minimum\ possible\ number\ of\ peaks)}{(maximum\ possible\ number\ of\ peaks\ -\ minimum\ possible\ number\ of\ peaks)}$$

A



B

Mixture Combinations	Σ Homozygote Counts	ΣA-A 1 bp Difference	Min/Max # of Expected Alleles	Σ Expected Alleles	ASR
C_{27}, C_{32}, C_{42}	9	0	2/5	70	0.63

Fig. 1. (A) An illustration of one expected/theoretical 3P mixture profile out of 19,600 different mixture genotype combination possibilities simulated from 50 single-source profiles found in the PROVEDIt dataset. The illustration was created in EuroForMix [3] using the genotypes of the following PROVEDIt single-source profiles (C_{27}, C_{32}, and C_{42}) where C refers to contributor. This output is not intended to reflect peak height variation, just the presence of ground truth genotypes. Different statistic metrics are shown at each of the autosomal STR loci. Homozygosity represents the count(s) of a homozygote genotype at a locus; A-A (1 bp) represents the instance(s) of a single base-pair difference between two alleles at a given locus; Number of alleles is the count of alleles expected to be observed; Expected alleles are alleles that are expected to be observed. (B) Summary statistics across the simulated mixture combination (N=21 loci) and the ASR.

3.3. Choosing a validation experimental design

This functionality is still under development and will aid the user in choosing combinations. After generating the complexity metrics, the user can then decide on and input into the software other levels of factor space desired to be covered, such as:

— Experimental NoC
— Apparent NoC (e.g., a 4P mixture that could be interpreted as 3P or 2P mixture)
— DNA quality (pristine or degraded)
— Total template amounts
— Level of A-A 1bp
— Level of allele sharing
— Contributors' template amounts or mixture ratios
— Number of runs per NoC value.

Using the statistical theory of factorial and fractional factorial experimental designs [7], the software will output candidate experimental plans based on user specifications to ensure reasonable coverage of the factor space.

3.4. Mixture Calculations

This functionality will aid in accurately calculating how much of each contributor's DNA is required to make the final volume of a desired concentration that contains the desired mixture ratios. The software will take the user's requirements (e.g., contributor's concentration, desired mixture ratios) and constraints (e.g., minimum pipetting amounts, DNA mass in the PCR reaction, minimum mixture stock solution) and provide the optimal strategy for making the desired mixtures.

4. Discussion

The scope of this research project was guided by discussions on validation challenges with forensic practitioners from different laboratories. Most forensic laboratories interpret their data using traditional methods (e.g., manual data formatting and entry in Microsoft Excel spreadsheets) or outsource their validation results to the manufacturers. In the latter case, labs still reanalyze a subset of the outsourced data to ensure independent verification, accuracy, reproducibility, and consistency with the analysis completed by the manufacturers. These approaches can be error-prone, time-consuming, and pose a significant delay in applying newer methods to casework. The GUI described here will be an accessible and easy-to-use inter-

face. It will be developed specifically to aid forensic practitioners in automating the design and interpretation of validation experiments without having to be an expert in programming languages or statistics. To enhance accessibility, the tool will be accompanied by instructions describing the installation of the software as well as tutorials and on-demand videos demonstrating the use of the tool.

5. Conclusion

In this work, we present the initial stages of the development of a standalone (i.e., not running over a network) and easy-to-use GUI. The GUI is designed to be integrated into the validation workflows and will assist the practitioners in automating (1) the design of validation samples and (2) the interpretation and visualization of the resulting data. This will free the analysts from the technicalities of manual data formatting and processing and allow them more time to explore, visualize, and understand the data output. The tool will also limit human errors, prevent data formatting issues, and increase the speed, effectiveness, and accuracy of analyzing data from validation studies.

6. Funding and Disclaimer

This work was supported by the Accelerating Forensic Innovation for Impact through the NIST Special Programs Office: *Forensic Genetics*. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. Certain commercial software, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose. All work involving NIST samples has been reviewed and approved by the NIST Research Protections Office.

7. Acknowledgments

We thank the U.S. forensic labs (CalDoJ, Jefferson County Regional Crime Laboratory, ATF, and Los Angeles County Sheriff's Department) for their valuable input and feedback on the software functionalities.

8. Conflict of interest statement

The authors declare no conflict of interest.

9. References

[1] GeneMapper™ ID-X Software v1.6, 2019.

[2] M.-H. Lin, S.-I. Lee, X. Zhang, L. Russell, H. Kelly, K. Cheng, S. Cooper, R. Wivell, Z. Kerr, J. Morawitz, J.-A. Bright, Developmental validation of FaSTR™ DNA: Software for the analysis of forensic DNA profiles, Forensic Science International: Reports 3 (2021) 100217.

[3] Ø. Bleka, G. Storvik, P. Gill, EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts, Forensic Science International: Genetics 21 (2016) 35-44.

[4] D. Taylor, J.-A. Bright, J. Buckleton, The interpretation of single source and mixed DNA profiles, Forensic Science International: Genetics 7(5) (2013) 516-528.

[5] L.E. Alfonse, A.D. Garrett, D.S. Lun, K.R. Duffy, C.M. Grgicak, A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt, Forensic Science International: Genetics 32 (2018) 62-70.

[6] P.C. Lynch, R.W. Cotton, Determination of the possible number of genotypes which can contribute to DNA mixtures: Non-computer assisted deconvolution should not be attempted for greater than two person mixtures, Forensic Sci Int Genet 37 (2018) 235-240.

[7] W.S. Conner, M. Zelen, Fractional Factorial Experiment Designs for Factors at Three Levels (National Bureau of Standards, Applied Mathematics Series, Vol. 54). U.S. Government Printing Office: Washington, DC (1959).

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 985-990
DOI: https://dx.doi.org/10.15304/cc.2025.1869

The impact of drop-out in the evaluation of the likelihood ratio with the continuous approach

Alessandra Caglià[1*], Manuela Saccente[1,2], Alessandra La Rosa[1], Stefano D'Amelio[2]

[1]*Forensic Science Police Service, Forensic Science Police and Cyber Security Central Directorate, Ministry of Interior- Rome, Italy.*

[2]*Department of Public Health and Infectious Disease, Sapienza University of Rome, Italy*

** Corresponding author at: Forensic Science Police Service-Via Tuscolana 1548, 00173-Rome, Italy*
E-mail address: alessandracaglia@gmail.com (A.Caglià), Phone:+390646543618

Abstract

As part of our internal validation of mixed genetic profiles and related bio-statistical calculations, we focused on one of the main aspects linked to the validation process, represented by the loss of genetic information, trying to identify the limits in the application of a probabilistic genotyping software (PGS).

We studied twelve mixtures of two contributors (2PM), simulating a progressive loss of genetic information with 0, 6, 12, 21 drop-out, corresponding to 0%, 15%, 30% and 50% loss of the person of interest (POI) genetic profile.

We calculated the likelihood ratio (LR), using the continuous probabilistic genotyping software EuroForMix (EFM), for the basic hypothesis [LR_{Base}: H1=POI+U/H2=U+U] and three LRs_{Fam}, replacing a relative of the POI (child/father or sibilings or cousin) in H2.

The expected decreases in the LR value, as the number of drop-out increased, made it possible to observe that there is no chance to determine a relative threshold value in the analyzed data, below which the LR is considered unreliable. On the other hand, this study suggested to us the possibility to use the LRs_{Fam} near to neutral value as a warning in the report for the court, especially in cases where the LR_{Base} is relatively low ($\leq 10^4$) and/or in those criminal cases such as burglary or robbery, where entire family clans can often be involved.

Keywords

Probabilistic genotyping software, DNA mixtures, EuroForMix, Drop-out

1. Introduction

The interpretation of complex genetic profiles, especially mixed profiles with more than two contributors and in which the Person of Interest (POI) suffers a partial loss of genetic information (drop-out), currently requires the approach of statistical evaluation of the analytical data and not just experiential skills of the forensic analyst.

Probabilistic genotyping software, used to calculate the likelihood ratio (LR) in the event of a comparison, raises the basic question "How can we be sure that a complex computer program is providing the right answer?" [1]

Unfortunately, there is no correct answer to this question, but there are various guidelines and reviews in the literature on the PGS validation issue, which can help a laboratory to test the reliability of the results [2].

During our internal validation of mixed genetic profiles and related bio-statistical calculations, we studied one of the main aspects linked to validation represented by the loss of genetic information, trying to identify the limits in the application of PGS.

So we asked ourselves: to what extent and in what ways can the occurrence of the loss of genetic information relating to the POI impact on the calculation of LR and its reliability? Is there a maximum number of drop-out beyond which the LR value may not be reliable?

2. Material studied, methods, techniques

The continuous probabilistic genotyping software used for this study was EuroForMix (vers. 3.3.4/4.0.8) [3], applying the Italian population allele frequencies and a theta factor equal to 0.01.

The GlobalFiler kit (21 autosomal loci) and the 3500 CE (run validated parameters) were used. The selected model was the one automatically chosen by EFM using the Optimal quantitative LR function (automatic mode search).

The LR value reported in the study is the log10LRML, i.e. the log10 of the maximum likelihood-based LR.

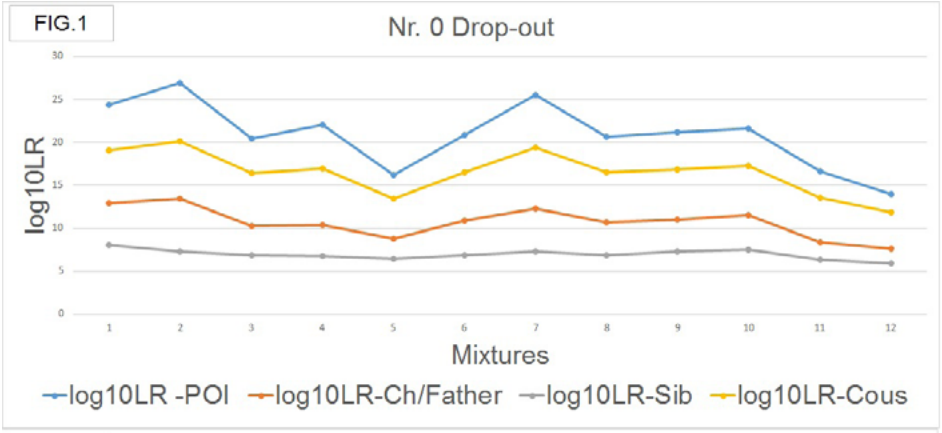
We selected two mixed genetic profiles of 2 contributors (Mixture Proportion (MX)~1:9) from our criminal cases routine, with reasonably known ground truth, and simulated a progressive loss of genetic information (from 1 to 21 sequential drop-out), up to loss 50% of the minor contributor (POI).

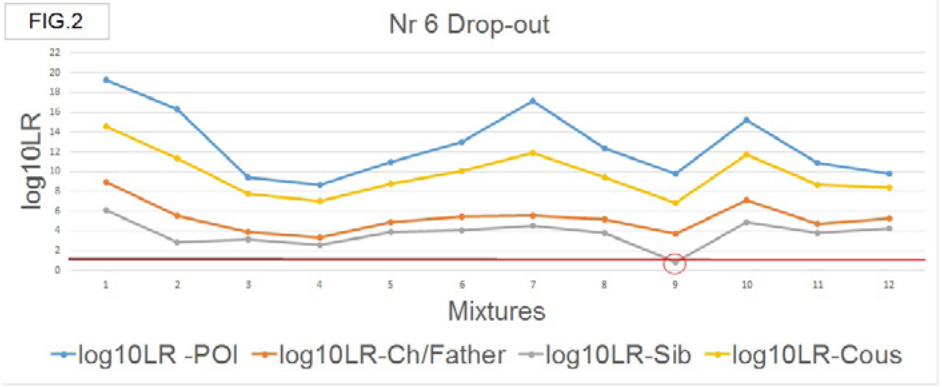
The choice of the dropped allele was made initially by removing the peaks with the lowest height and the highest molecular weight (mimicking as much as possible what generally happens in real cases if stochastic effects occur).

Then we calculated the likelihood ratio for the basic hypothesis [LRBase: H1=POI+U/H2=U+U] and three LRsFam, replacing a relative of the POI (child/father or sibilings or cousin) in H2.

We also repeated the study on ten 2PM with major/minor contributors from the PROVEDIt dataset [4] (with various MX 1:2, 1:4 and 1:9, degradation and DNA input), studying these profiles simulating a progressive loss of genetic information with 0, 6, 12, 21 drop-out, corresponding to 0%, 15%, 30% and 50% loss of POI genetic profile (see Fig. 1-4).

It should be noted that EFM calculates a weight of evidence for all the various genotypic combinations, even particularly unlikely ones (in these cases other software can evaluate specific locus LR =0), always returning a total LR value.





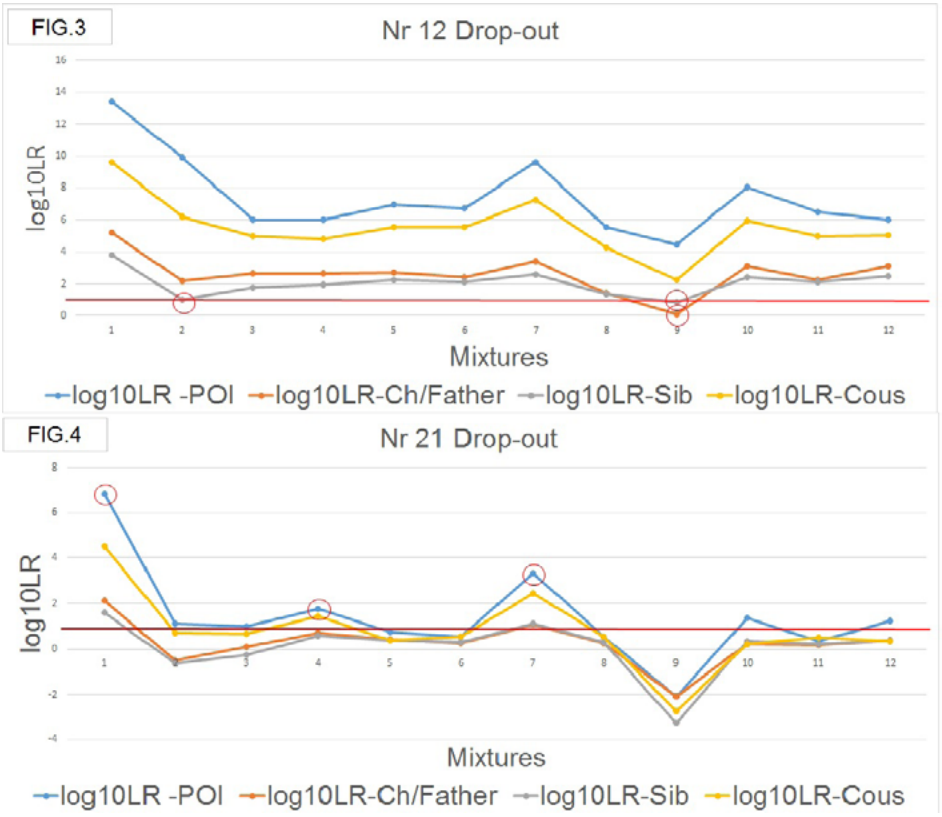


Fig 1-4: The LR_{Base} (log10LR-POI the blue line) and LR_{Fam} (log10LR-Ch/Father the orange line; log10LR-Sib the gray line; log10LR-Cous the yellow line) values studied in twelve 2PM.

3. Results

In the context of testing the limits of application in the use of a PGS, identifying a possible threshold of acceptability of the maximum number of drop-out, is linked to the issue of the risk related to the exclusion of a true contributor or the inclusion of a false contributor (type 1 and 2 errors [5]).

The expected decrease rate in the LR_{Base} and LR_{Fam} values, as the number of drop-out increased (see Fig.1-4), is quite variable in the mixtures studied, and depends among other factors on the dropped alleles (more or less rare) and consequently on the allele frequencies of those remaining in the profile (data not shown in the present work).

As expected, the greatest decrease in the LRs_{Fam} occurs in the hypothesis in which siblings are present in H2 (see gray lines log_{10}LR-Sib in Fig.1-4), while the LRs_{Fam} related to the presence of a POI-Cousin in H2 are the closest to the basic hypothesis (see yellow lines log_{10}LR-Cous in Fig.1-4).

It was possible to observe how with 6 drop-out (15% loss of genetic information), the LR_{Base} values are between 8.66 and 19.24, and only in one case the LR_{Fam} <1 (see Fig.2- red circle Mixture No.9).

In the case of 12 drop-out (30% loss), the LR_{Base} values are between 4.44 and 13.39, and only on three occasions the LR_{Fam} is ≤1 (see Fig.3-red circles).

Finally, with 21 drop-out (50% loss), it is possible to observe that the LRs_{Fam} are always less than 1, with the exception of three mixtures (see Fig.4- red circles), where in two of these, it is possible to underline an unexpected high LR_{Base} (6.84 and 3.33, respectively), despite such a high number of drop-out.

4. Discussion

The loss of information does not affect the calculation of the LR in the same way in all studied mixtures. Furthermore it can change in relation to various factors such as, the number of contributors (NOC), the MR and whether the POI is the major or minor contributor, the quality of the profile (degradation, LT-DNA), the allele frequencies of the dropped alleles and those relating to the remaining alleles, the degree of overlapping.

In relation to the reliability of low LR values, some laboratories use a threshold value (generally 104) to report the LR, below which the LR value is considered inconclusive. The analyses of these results therefore suggested to us the possibility to relate the LRsFam in the report for the court, if they are close to the neutral value (0<log10LR>1) or veer in favor of H2 support (log10LR<0).

These LRFam values could be considered as a warning, especially in cases where the LRBase is relatively low (≤ 104) and/or in those criminal cases such as burglary or robbery, where entire family clans can often be involved. In these cases, EuroForMix offers, in addition, a further opportunity of performing a non-contributor test.

All these considerations are important in order to try to find the most objective, impartial and transparent approach to explaining the meaning of the LR in the court, where the final users are represented by legal people who are not normally experts in forensic statistical evaluation of the weight of evidence.

5. Conflict of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6. References

[1] P.Gill, *Ø. Bleka, O. Hansson et al. Forensic practitioner's Guide to the interpretation of complex DNA profiles, Accademic Press, 2020. p.277*

[2] M.D. Coble, J. Buckleton, J.M. Butler, et al., DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications. Forensic Sci. Int. Genetics.25 (2016) 191-197.

[3] Ø. Bleka, G. Storvik, P Gill, EuroForMix: an open source software based on a continuos model to evaluate STR DNA profiles from a mixture of contributors with artefacts. Forensic Sci Int Genetics. 21(2016) 35-44.

[4] E.L. Alfonse, A.D. Garrett, D.S. Lunb et al., A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt, Forensic Sci. Int. Genetics. 32 (2018) 62-70.

[5] C.C..Benschop, A. Nijveld, F.E. Duijs et al., An assessment of the performance of the probabilistic genotyping software EuroForMix: Trends in likelihood ratios and analysis of Type I and II errors, Forensic Sci Int Genetics. 42 (2019) 31-38

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 991-997
DOI: https://dx.doi.org/10.15304/cc.2025.1869

An overlooked treasure; the relationship between the back stutter rate and the double back stutter rate

Minoru Asogawa

Machida, Tokyo, Japan

** Corresponding author at: cmumino@gmail.com,*

Abstract
Stutter, which is an artifact of PCR and prevent correct mixture analysis, has become an important research issue. MPS experiments have enabled the precise observation of the stutter amplicons. The average back stutter ratio () and the average double back stutter ratio () vary depending on the sample and experimental conditions used. However, a consistent finding was that can be expressed in the form of a quadratic equation of. To elucidate this quadratic expression, we undertook a modelling exercise. In this modelling, we considered conventional SMM and its extendions, and demonstrated that can be expressed in the form of a quadratic equation of. Utilising this relationship, we demonstrate that can be explained as a quadratic equation for the number of repeats, which is consistent with the experimental data. Furthermore, we suggest that this model can be applied to MSI data.

Keywords
Back stutter ratio, double back stutter ratio, SMM, MSI

1. Introduction

In the experiments utilizing MPS, the precise observation of the number of backstutter amplicons and double backstutter amplicons was observed. The mean values of the back stutter ratio () and the double back stutter ratio () exhibit variability, yet a consistent finding is that can be expressed in the form of a quadratic equation of. However, no model existed to explain this phenomenon. Consequently, we considered the conventional SMM and extended models and demonstrated that can be expressed in the form of a quadratic equation of. Utilising this relationship,

we established thatn can be explained as a quadratic expression for the number of iterations, which is consistent with the experimental data.

2. Material studied, methods, techniques, and Results

In the supplementary of Ran Li's paper [1] contains the stuttering data (N-4, N-3, N-2, N0, N+1, and N+2 variants), which is obtained by analyzing the ForenSeq DNA Signature Prep Kit on a MiSeq FGx instrument. The data is the result of 750 peripheral blood samples from human subjects including the stutters of 58 STRs. By using this data, the relationship between the back stutter rate and the double back stutter rate is shown in Fig. 1, which shows that all the locus confirmed that (orange line).

In the supplementary of Hoogenboom's paper contains the stuttering data (vs. repeat length), analyzed via the Promega Powerseq Auto System on a MiSeq instrument [2]. The data is analysis result of 450 samples from human subjects. In the paper the relationship between the back stutter rate and the double back stutter rate versus repeat length of D22S1045 are dawned (Fig. 2). In Fig. 2 each data point on the graph represents for the blue and for the orange. As illustrated in Fig. 2, can be approximated by a linear equation (depicted as a black solid line in the figure) for the number of iterations (). Furthermore, the minimum number of iterations () can be considered in relation to the occurrence of back stutter. The left part of Fig. 2 is the enlarged graph in the vertical axis. The black curve represents the data presented in the original quadratics curve [2].

3. Discussion

3.1. Using SMM

SMM employs a multinomial distribution as the probability of no increase (), increase without stutter (), and increase with stutter () at each step of the PCR processes [3]. Utilizing these probability parameters, the number of templates () and the number of PCR cycles (), the mean () of the number of unstuttered amplicons count () is estimated by. Similarly, the mean of the back stutter amplicon count () can be estimated by. This equation can be explained by the fact that the change from "no stutter" to "back stutter" occurs only once in each PCR process, and the number of such cases is times. It is possible to utilize more detailed simulations

differentiating between templates, semi-templates, and amplicons and consider the probability of PCR amplification. In the interest of providing a comprehensive overview, this study does not distinguish between these amplified products.

In the SMM model, the double back stutter is produced through two consecutive stuttering processes; from "no stutter" to "back stutter" and from "back stutter" to "double back stutter", the number of such cases is times.

Therefore, the average of the double back stutter amplicon count () can be estimated by, where represents the probability of back-stutter production, which may be slightly smaller than. The is quadratically increasing with respect to. If we approximate,, then can be approximated as. This approximation has not been published, to the best of our knowledge. In the paper by Jos [3] the probabilities of PCR on the template and amplicon are distinguished. Assuming that these probabilities are equal (), comparable results can be derived for the estimation of and [3]. From the previous equations, if and are quantified in the experimental setting, the value of can be estimated. If we assume the mean of the back stutter ratio () is approximated by, then. Note that this relation is independent to and repeat structure. This theoretical equation has been demonstrated in prior research utilizing SMM [3,4]. As will be discussed subsequently, the approximation of with is not mathematically appropriate approach. Nevertheless, the simulation results indicate that this is a highly accurate estimation. If we approximate the mean of the double back stutter ratio (), then we can express it as. This is equivalent to:. When is approximated to zero and is approximated to, can be approximated to. Similarly, can be approximate to.

3.2. Approximateby

As previously stated, approximating by is not a mathematically appropriate approach. This is because the back stutter ratio () is the observation for each experiment as "the amount of back stutter amplicon" divided by "the non-stutter amplicon." The average of this value over all trials represents the average (). In this context, and represent and, respectively. Consequently, and the are not mathematically identical. We developed code with R, similar to PCRsim, and performed 100,000 simulations with =322[copy], =0.95, and =0.012, under the condition of 20 PCR cycles. Through the simulation it is shown that is a good estimator of. Since the is given by ratio statistics, when both non-stutter amplicon and back stutter distribution is approximated by a LogNormal distribution, then the back stutter ratio can be approximated by a LogNormal distribution.

3.3. Saturation is currently under consideration by SMM

In SMM, the value of is assumed to increase as the number of PCR cycles,, is increased. In actual PCR, the value of does not increase by a greater amount than a constant number even if the PCR cycle number, increases above a certain level. This is due to the occurrence of saturation, which results in the establishment of a plateau. It may therefore be anticipated that will also decline to a constant value. To simulate this phenomenon, a saturation-effect SMM was constructed using the R language. Firstly, to ascertain the quantity of total amplicons in a saturated state, 20 PCRs were conducted with =322[copy], =0.95 and =0.012, with the assumption that no saturation effect was present. The mean of all amplicons at that time was then taken as the saturated amplicon amount (). For the sake of simplicity, we will assume that. In the simulation with saturation effects, the ratio of this to the total amplicon volume after each frequency was employed to as the PCR efficiency () at each cycle. This value is then multiplied by all probability. In the beginning of PCR, $\alpha \cong 1.0$. However, as the total amplicons increase exponentially and approaching the ratio approaches to 0. In the simulations, 35 cycles were performed with saturation effects incorporated. The simulation result indicates that m=20 without saturation effects, is the good estimator for after 35 cycles with saturation effects.

3.4. The model permits the implementation of SMM{0,-1,-2} stutter within a single PCR process

As demonstrated in the preceding chapter, the SMM allows for the estimation of as. In the published MPS data, is approximately equal to. Thus experimental is larger than the SMM estimation. To explain this difference, in addition to the SMM, a model was constructed which incorporates a "double back stutter" process during one time in one PCR cycle. Assuming that the probability () of this process increasing is proportional to the number of PCR processes, the value of is modified as follows:

In the absence of detailed knowledge regarding the specifics of the experiment condition, we may nonetheless make certain assumptions. For instance, we may posit that and than. Because, the second term represents the, which should be equal to. This implies.

3.5. The model permits the implementation of {0,-1,-2} stutter within a single PCR process

Fig. 2 depicts a graph of the number of repeats () versus, versus for D22S1045. In the SMM, it can be approximated as is approximately, as illustrated by the blue

line. The blue line is smaller than the actual double back stutter rate. The addition of the above 'probability of increase with double back stutter ()' results in a green line. Here, the 'probability of increase with double back stutter ()' is assumed to be proportional to (). The addition of this effect results in the orange line, which is in good agreement with the values presented in the paper (black solid line).

3.6. The model permits the incorporation of a stutter of up to ±{0,1,2} stutter in a single PCR process.

An additional extension of the aforementioned model, which incorporates forward stutter and double forward stutter rates, is illustrated in Fig. 3. For the sake of simplicity, the figure depicts for back stutter as 2 and for forward stutter as 2. The distribution of stutters after one PCR process, when there are [number] of 5 repates (rightmost column vector), is represented as the column vector. is the PCR efficiency. The leftmost column vector represents the average amplicon distribution after one cycle. Note that the can be measured by using real-time PCR experiments. For the sake of simplicity, we have restricted ourselves to the range of possible stutters stutter, but it is assumed that an extension of more than 2 is necessary when the number of bases in the repeats is large. By using the experimental data of stutter distribution, it is possible to estimate these parameters.

4. Conclusion

We were able to explain the observed in the MPS data using a model that extends the SMM. We also showed that can be expressed as a quadratic for the number of repeats. Since the model can estimate PCR artifacts, it is expected to be applicable to the analysis of mixed samples and to the analysis of microsatellite instability (MSI) data related to the diagnosis of cancer.

5. Acknowledgments

The author would like to thank to Dr. J. Hoogenboom and Dr. Ran Li, for their valuable generous discussion and supplying the MPS data.

6. Conflict of interest statement

None

7. References

[1] Li R, et. al., Characterizing stutter variants in forensic STRs with massively parallel sequencing, Forensic Sci Int Genet. 45:102225, (2020).

[2] Jerry Hoogenboom, et. al., FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise, Forensic Sci Int Genet. Mar:27:27-40., (2017).

[3] Jos Weusten, et. al, A stochastic model of the processes in PCR based amplification of STR DNA in forensic applications, Forensic Sci Int Genet. Jan;6(1):17-25, (2012).

[4] Soren B Vilsen et. al., Stutter analysis of complex STR MPS data, Forensic Sci Int Genet. Jul:35:107-112, (2018).

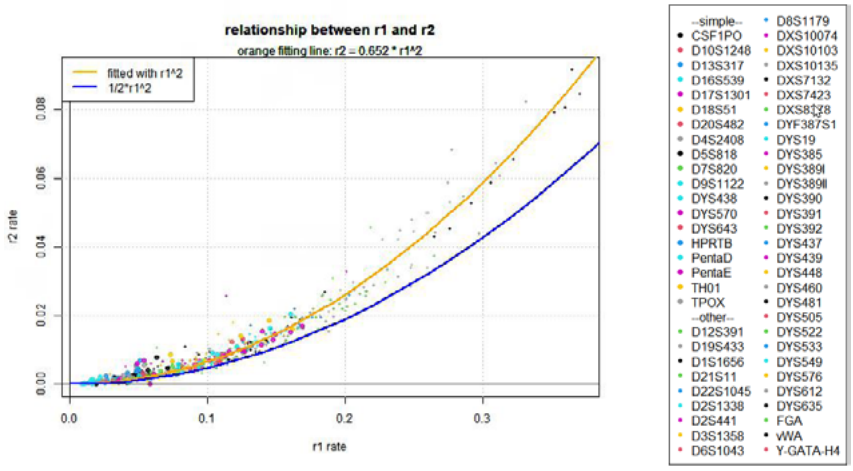


Figure 1. Relationship between r1 and r2 for 58 loci. shows that all the locus confirmed that (orange line). The blue line is the theoretical from SMM.

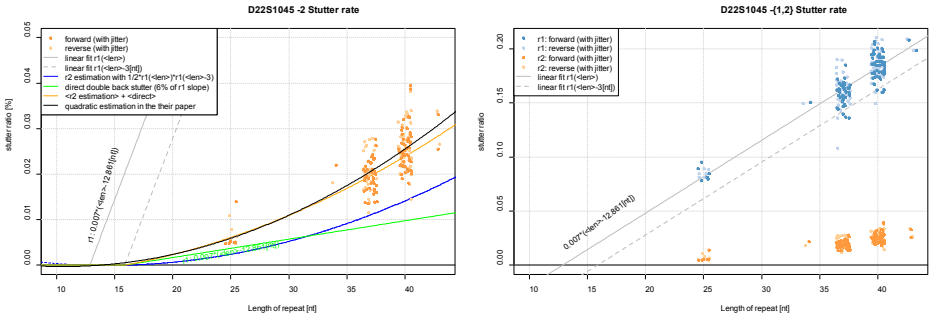
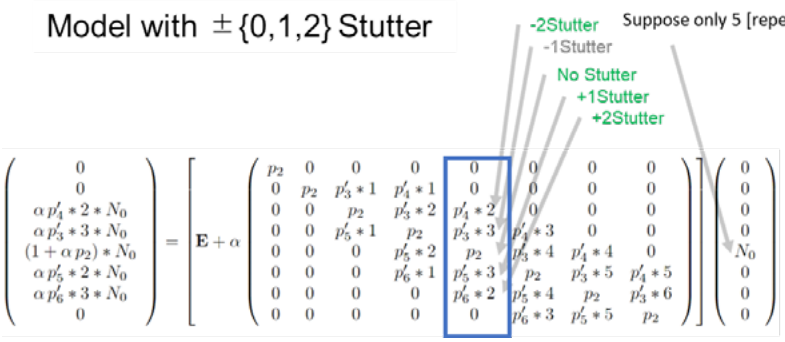


Figure 2. The relationship between the back stutter rate and the double back stutter rate versus repeat length of D22S1045. each data point on the graph represents for the blue and for the orange. As illustrated in Fig. 2, can be approximated by a linear equation (depicted as a black solid line in the figure) for the number of iterations (). Furthermore, the minimum number of iterations () can be considered in relation to the occurrence of back stutter. The left part of Fig. 2 the rate on the vertical axis has been expanded. The black curve represents the data presented in Hoogenboom's original quadratics curve presented in the paper [2].



$$\begin{pmatrix} 0 \\ 0 \\ \alpha\, p'_4 * 2 * N_0 \\ \alpha\, p'_3 * 3 * N_0 \\ (1+\alpha\, p_2) * N_0 \\ \alpha\, p'_5 * 2 * N_0 \\ \alpha\, p'_6 * 3 * N_0 \\ 0 \end{pmatrix} = \left[\mathbf{E} + \alpha \begin{pmatrix} p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_2 & p'_3 * 1 & p'_4 * 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_2 & p'_3 * 2 & p'_4 * 2 & 0 & 0 & 0 \\ 0 & 0 & p'_5 * 1 & p_2 & p'_3 * 3 & p'_4 * 3 & 0 & 0 \\ 0 & 0 & 0 & p'_5 * 2 & p_2 & p'_3 * 4 & p'_4 * 4 & 0 \\ 0 & 0 & 0 & p'_6 * 1 & p'_5 * 3 & p_2 & p'_3 * 5 & p'_4 * 5 \\ 0 & 0 & 0 & 0 & p'_6 * 2 & p'_5 * 4 & p_2 & p'_3 * 6 \\ 0 & 0 & 0 & 0 & 0 & p'_6 * 3 & p'_5 * 5 & p_2 \end{pmatrix} \right] \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ N_0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Figure 3. He model permits the incorporation of a stutter of up to ±{0,1,2} stutter in a single PCR process. The rightmost column vector indicates the amplicon distribution before PCR cycle. For the sake of simplicity, it is depicted that there are [number] of 5 repeats. The leftmost column vector represents the average amplicon distribution after one cycle. In the amplification matrix, for back stutter is 2 and for forward stutter is 2. The leftmost column vector represents the average amplicon distribution after one cycle.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 999-1003
DOI: https://dx.doi.org/10.15304/cc.2025.1869

DNA mixture interpretation: A guiding principle to identify the best supported propositions

Klaas Slooten[1], Vincent van Marion[1], Jord H.A. Nagel[1], Charissa van Kooten[1], Corina C.G. Benschop[1*]

[1]*Netherlands Forensic Institute, The Hague, The Netherlands.*

** Corresponding author at: Netherlands Forensic Institute, c.benschop@nfi.nl*

Abstract
Substantial statistical and computational advances necessitate revision of guidelines for mixture interpretation in DNA casework. Our extended procedure for DNA mixture interpretation is centered around a guiding principle: Identifying the proposition that is best supported by the data within the context of the case. Examples of the application of this guiding principle are presented.

Keywords
Forensic science, DNA mixtures, likelihood ratio, probabilistic genotyping software, interpretation guidelines, exhaustive propositions.

1. Introduction

Traditional DNA mixture interpretation guidelines [1], written for comparisons between trace profiles and reference profiles of persons of interest (PoIs), follow a fixed stepwise approach with definite decisions being made. E.g., setting the number of contributors at a fixed number. In our laboratory, these rigid guidelines have been followed for a long time. However, technological developments as well as statistical insights have prompted us to revise and extend our guidelines for mixture interpretation in DNA casework. Here, we share the philosophy behind our procedure for DNA mixture interpretation as currently employed within our laboratory, which is centered around a guiding principle: Identifying the propositions that are best supported by the data within the context of the case (Fig. 1).

This principle is versatile and can be applied across all steps of DNA profile interpretation. By logical reasoning, the reporting officer can determine whether weight of evidence calculations are needed, identify relevant explanations for the DNA profile, and decide how to report the evidence. Unlike traditional methods that limit the interpretation to a predefined pair of propositions (i.e. hypotheses), this approach allows for a dynamic evaluation process. Propositions can be customized and refined based on insights from the statistical analysis, resulting in a more complete assessment of the DNA evidence. This versatile procedure has been implemented in our daily casework at the NFI using DNAStatistX [2] and based on the methodology as described in [3].

2. Traditional versus extended procedure for DNA mixture interpretation

Fig. 1 and Table 1 outline the main differences between the traditional and extended procedure for DNA profile interpretation.

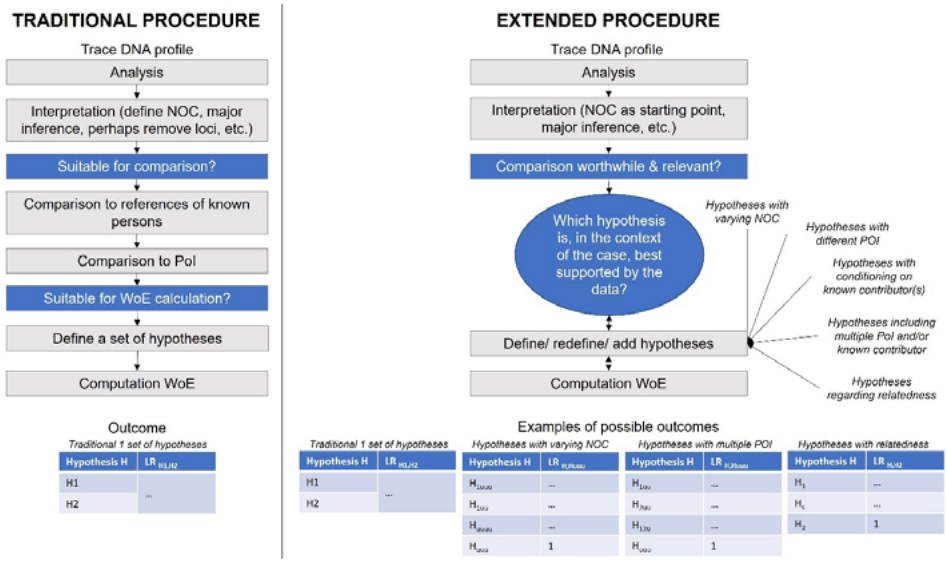


Figure 1. Overview of the traditional (left) step-wise procedure to DNA profile interpretation and the extended, flexible, procedure (right) in which the interpretation is centered around the question "Which hypothesis is, within the context of the case, best supported by the data?".

Table 1. Differences between the traditional and extended procedure for DNA mixture interpretation.

	Traditional procedure	Extended procedure
Suitability for comparison	Judged by expert prior to comparison	All profiles may be subjected to comparison. Not all may be worthwhile or relevant for statistical evaluation. Validity of evaluation is judged (see Validation).
Choice of hypotheses	Expert defines set of hypotheses, fixed throughout evaluation	Relevant propositions are defined as a starting point. More propositions may be added when expected to be relevant and potentially discriminatory. They may alter conclusions or widen the context in which conclusions apply.
Number of contributors (NoC)	NoC is defined prior to comparison and kept fixed	Initial assessment of the NoC as a starting point. Additional hypotheses with varying NoC can be evaluated along the interpretation process.
Results	Two likelihoods, one likelihood ratio (LR)	A likelihood of each hypothesis, evidential weight summarized by one or more LRs.
Validation of results	Expert opinion	Expert opinion aided by statistical model validation.

3. Application in casework: Two examples

In case of a single PoI, certainty about the NoC, and no relatedness to query, the revised procedure amounts to the original. I.e., calculating likelihood ratios (LRs) for the two hypotheses according to which the PoI did (resp. did not) contribute. Tables 2 and 3 present two more complex casework examples where the extended procedure for DNA mixture interpretation has been particularly useful. After an initial assessment of the (mixture) DNA profile, possible propositions are listed ('exhaustive' propositions). Next, LRs are calculated using each hypothesis against a default alternative. From the results one can assess the evidence per PoI. If there is no support for inclusion of a PoI, the hypotheses including this PoI are omitted and the evidence per (remaining) PoI is re-assessed. A helpful excel sheet for this combined evaluation is available upon request and incorporation of the procedure is listed for a future version of DNAxs/ DNAStatistX.

Table 2. Overview of assessment of casework example 1, with uncertainty on the NoC.

A. *List all hypotheses*

Number	Investigated hypotheses H	$LR_{H,Huuu}$
1	H_{12uu}	10^{20}
2	H_{1uuu}	$10^{10.5}$
3	H_{2uuu}	$10^{9.5}$
4	H_{uuuu}	$10^{-0.5}$
5	H_{12u}	10^{-3}
6	H_{1uu}	10^{11}
7	H_{2uu}	10^{10}
8	H_{uuu} [a]	1

[a] Default alternative chosen in this case

C. *Define best supported hypotheses and write conclusions*

Reported conclusion

Eight hypotheses were considered, corresponding to whether each of the PoI is assumed to be a contributor or a non-contributor, and assuming that three or four individuals contributed DNA to the trace. The contributors other than the PoI were considered to be unknown individuals without relatedness to each other or the PoI. A statistical analysis showed that the trace profile is at least a billion times more likely if both PoI together with two unknown individuals contributed, than if any of the seven other considered hypotheses is true.

B. *Assess evidence per PoI*

PoI	H1	H2	LR	Support for contribution?
PoI 1	$10^{20} + 10^{10.5} + 10^{-3} + 10^{11}$	$10^{9.5} + 10^{-0.5} + 10^{10} + 1$	Appr. $10^{20-10} = 10^{10}$	Yes
PoI 2	$10^{20} + 10^{9.5} + 10^{-3} + 10^{10}$	$10^{10.5} + 10^{-0.5} + 10^{11} + 1$	Appr. $10^{20-11} = 10^{9}$	Yes

Table 3. Overview of assessment of casework example 2. Three PoI. Evidence for presence of DNA of PoI 1&2, but not for PoI 3. Best supported hypothesis is presence of PoI 1, PoI 2 and two unknowns.

A. *List all hypotheses*

Hypothesis number	Investigated hypotheses H	$LR_{H,Huuu}$
1	H_{123u}	$10^{13.2}$
2	H_{12uu}	$10^{14.7}$
3	H_{13uu}	$10^{8.7}$
4	H_{23uu}	$10^{10.3}$
5	H_{1uuu}	$10^{8.2}$
6	H_{2uuu}	$10^{8.7}$
7	H_{3uuu}	$10^{5.8}$
8	H_{uuuu}	1

B. *Assess evidence per PoI*

PoI	H1	H2	LR	Support for contribution?
PoI 1	$10^{13.2} + 10^{14.7} + 10^{8.7} + 10^{8.2}$	$10^{10.3} + 10^{8.7} + 10^{5.8} + 1$	Appr. $10^{14.7-10.3} = 10^{4.4}$	Yes
PoI 2	$10^{13.2} + 10^{14.7} + 10^{10.3} + 10^{8.7}$	$10^{8.7} + 10^{8.2} + 10^{5.8} + 1$	Appr. $10^{14.7-8,7} = 10^{6}$	Yes
PoI 3	$10^{13.2} + 10^{8.7} + 10^{10.3} + 10^{5.8}$	$10^{14.7} + 10^{8.2} + 10^{8.7} + 1$	Appr. $10^{13.2-14.7} = 10^{-1.5}$ = appr 1/30	No

C. *Omit hypotheses for PoI without support for contribution*

	PoI 1	PoI 2	PoI 3	Hypotheses numbers
Retain hypotheses	Yes	Yes	No	2, 5, 6, 8
Omit hypotheses	No	No	Yes	1, 3, 4, 7

D. *Re-assess evidence per PoI*

PoI	H1	H2	LR	Support for contribution?
PoI 1	$10^{14.7} + 10^{8.2}$	$10^{8.7} + 1$	Appr. $10^{14.7-8.7} = 10^{6.0}$	Yes
PoI 2	$10^{14.7} + 10^{8.7}$	$10^{8.2} + 1$	Appr. $10^{14.7-8,2} = 10^{6.5}$	Yes

4. Concluding remarks

The procedure for mixture interpretation as outlined here aids experts in case of uncertainty on the NoC and/or in presence of multiple PoI, even if PoIs are related to each other. In such scenarios, this procedure is now consistently applied in our casework practice. The continuous advancement of probabilistic genotyping models further streamlines the evaluation of lists of hypotheses. The method clarifies in which context results are valid and prompts to reflect whether evaluation of additional propositions is potentially useful. This procedure follows international guidelines [4] and we believe that the presented procedure is the most natural extension of the LR framework from two to an arbitrary number of propositions.

5. Conflict of interest statement

Authors have no conflict of interest to declare.

6. References

[1] TM Clayton et al. Analysis and interpretation of mixed forensic stains using DNA STR profiling. Forensic Science International 91.1 (1998), pp. 55–70.

[2] C. Benschop et. al. DNAxs/DNAStatistX: Development and validation of a software suite for the data management and probabilistic interpretation of DNA profiles. Forensic Science International: Genetics 42 (2019), 767 pp. 81–89.

[3] K. Slooten. The comparison of DNA mixture profiles with multiple persons of interest. Forensic Science International: Genetics 56 (2022) 102592.

[4] P. Gill et al. DNA commission of the International society for forensic genetics: Assessing the value of forensic biological evidence-Guidelines highlighting the importance of propositions: Part I: evaluation of DNA profiling comparisons given (sub-) source propositions. Forensic Science International: Genetics 36 (2018), pp. 189–202.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1005-1011
DOI: https://dx.doi.org/10.15304/cc.2025.1869

MPS vs CE STR data, similarities and differences and the impact on LR calculations using probabilistic genotyping systems

Kristiaan J. van der Gaag[1,*,**], Oanelle Gléonec[1,**], Jerry Hoogenboom[1], Corina C.G. Benschop[1], Titia Sijen[1]

[1] *Division of Biological Traces, Netherlands Forensic Institute, The Hague, the Netherlands.*

** Corresponding author at: Division of Biological Traces, Netherlands Forensic Institute, Laan van Ypenburg 6, 2497 GB, The Hague, The Netherlands, k.van.der.gaag@nfi.nl*

*** These authors contributed equally.*

Abstract

The number of commercial kits for Massively Parallel Sequencing (MPS) of short tandem repeats (STRs) for human identification is rapidly increasing, and lab protocols are becoming easier and more suitable for high throughput use. While multiple studies have shown advantages of MPS STR sequencing compared to Capillary Electrophoresis (CE) analysis, only few studies have addressed the statistical likelihood ratio (LR) calculations on MPS STR data. While MPS and CE share many similar factors, this study addresses how the differences in the details of the data and the laboratory workflows impact the current CE-based statistical models when handling MPS data. Specifically, we focus on differences between replicates of the same mixture analyzed with different read depth in multiple MPS runs. LRs were calculated using DNAStatistX for complex 3- and 4-person DNA mixtures analyzed in triplicate using the IDseek® OmniSTR™ Global kit. Differences between the triplicate analyses often impacted the LR calculations resulting in failed validation of the used LR model. In this study, we discuss the different approaches that have been explored and provide recommendations for addressing MPS-specific factors impacting LR calculations.

Keywords

MPS, STR sequencing, LR calculation, replicates, normalization.

1. Introduction

MPS-based STR analysis can provide an increased discriminative power compared to CE-based analysis and use smaller amplicons [1,2]. Still, studies focusing on weight of evidence calculations for MPS data remain limited [3,4]. For LR calculations, *many factors are comparable such as balance, stochastic variations, and CE peak heights are similar to MPS reads counts. However, the respective workflows for CE and MPS may cause some essential differences specifically impacting LR calculations (*Figure 1*).*

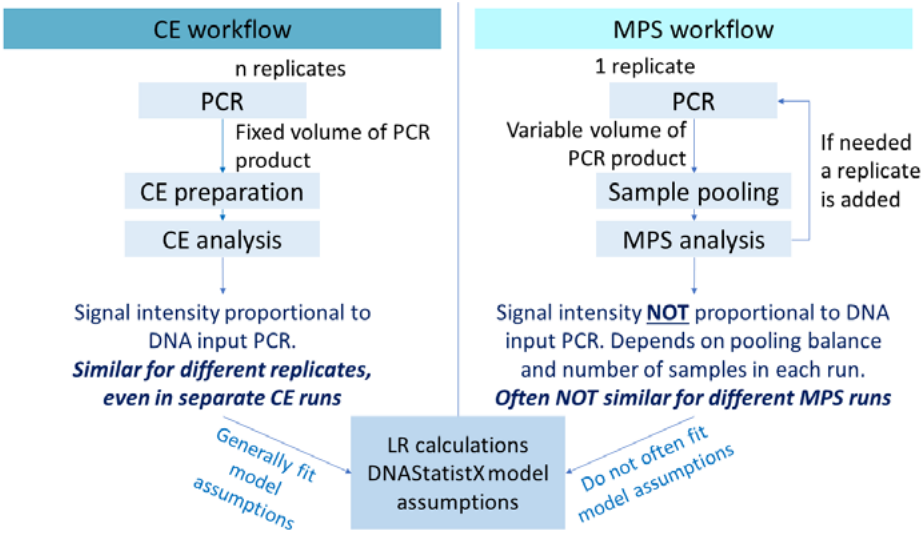


Figure 1. Difference between CE and MPS workflow impacting LR calculation using replicates in DNAStatistX, using the EuroForMix based LR model.

A key difference is the assumption that signal intensity is proportional to DNA input. While this holds true for CE, it is not necessarily the case for MPS, where signal intensity depends on pooling strategies both between and within runs. The more samples are pooled in a run, the fewer reads are available per sample, leading to between-run variation. For instance, a replicate analyzed in a run with fewer samples will generate more reads than one in a run with more samples. Within a run, lower DNA input results in more primer dimers utilizing reads, leaving fewer reads for genuine STR amplicons.

Additionally, in forensic casework, it is common to start with a single PCR and, based on performance and profile complexity, decide whether to perform additional replicates later. Therefore, variation in signal intensities between MPS replicates can be expected, which violates an assumption made by quantitative LR models such as EuroForMix used by DNAStatistX and leading to failed model validations [5,6]. The model validation within DNAStatistX and EuroForMix is an important quality check of whether the underlying gamma distribution model for peak heights (along with the defined propositions and parameter settings) explains the observed data well. I.e., if the cumulative probabilities for the observed read counts fit those for the model expected read counts. If there are too many signals not fitting the expectations, the model validation is scored as failed and require extra attention. It was advised that LRs are not reported if the model validation failure cannot be solved or explained.

To address the issue of failed model validations using MPS replicates data, we focused on the difference in intensity between replicates of mixed samples analyzed with IDseek® OmniSTR™ Global kit and its influence on model validation results for LR calculations. While it is possible to use a replicate independent LR model such as EFMrep [7], the approach described in this study proposes to normalize the read counts between replicates to allow the data to comply with the original model assumptions.

2. Material studied, methods, techniques

2.1. Material

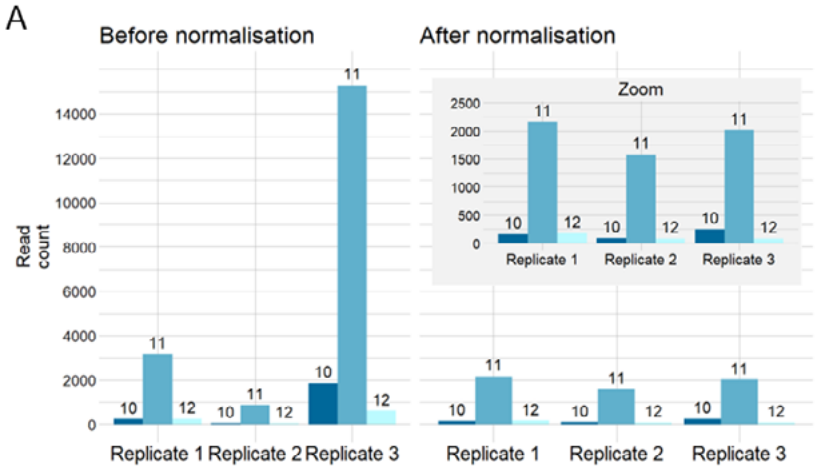
For this study, 36 complex mixtures with 3 and 4 contributors in different proportions and absolute DNA inputs from three donor sets were analyzed in triplicate with the IDseek® OmniSTR™ Global kit (NimaGen) on the MiSeq FGx (Qiagen) sequencing system. Library preparation was performed according to protocol except for pooling and pool purification (van der Gaag et al, submitted manuscript). Pooling was based on quantification by SYBRgreen qPCR [2]. Purification of the pool was performed using Ampure XP beads (Beckman Coulter) instead of Ampliclean beads (NimaGen) [2]. Data analysis was performed using FDSTools (v1.2.0) [8]. LR calculations were carried out using DNAStatistX (v2.1.1) [6].

2.2. Method

To evaluate the dispersion of signal intensity between replicates of a sample, the average squared deviation to the median is used. This measure increases when

there is more variation in total read count between replicates. This was compared to model validation results.

The proposed normalization approach focused on not altering allele proportions between alleles and markers within one replicate while balancing intensities between replicates. The goal of the normalization process is to equalize the total read count of replicates by applying a replicate-specific factor based on the value of the analytical threshold (to ensure that the lowest detected alleles remain above the analytical threshold) (Figure 2).



B

$$1.\ \frac{Total\ reads}{(Lowest\ reads/replicate/analytical\ threshold\ (here\ 15))}$$

Replicates	Total reads / replicate	Lowest reads / replicate	Max reads required for the allele with lowest reads to equal analytical threshold	Factor used
A	110384	22	75262	0.68
B	41417	21	29583	1.83
C	572923	237	36261	0.13

$$2.\ \frac{Total\ reads}{Max\ of\ minimum\ reads\ required}$$

Figure 2. (A) Effect of normalization on the signal intensity and allelic proportions of a marker from three MPS replicates. (B) Normalization process example with rounded values.

The factor is calculated by:

1. Determining in one replicate the lowest number of reads. This number is first divided by the value of the analytical threshold, and the total number of reads for the replicate is divided by this quotient. The resulting number is the total amount of reads required for the replicate. [Total reads/ (Lowest reads per replicate/Analytical threshold)]. This step is repeated for each replicate of a sample.
2. The replicate-specific factor is then calculated by dividing the total number of reads by the highest total amount of reads required from the previous step [Total reads/Max of minimum reads required].

The replicate-specific factor is then multiplied by the signal intensities per replicate.

3. Results / Discussion

The variation in total read count between replicates of 36 samples with failed and passed model validation is available in Figure 3. The average squared deviation to the median is higher for samples with failed model validations. Beside intensity imbalances between replicates, some analyses with failed model validations showed low read counts and artifacts such as drop-in and drop-out alleles. After normalization, six out of the seven samples with failed model validation (including all with intensity imbalances between replicates) resulted in a passed model validation and LRs remained similar.

In instances where the model does not fit after normalization, other aspects may cause the model validation failure. Such as, too many peaks to be explained by the number of contributors that are assumed under the propositions. Checking the data and perhaps considering alternative propositions may solve the model validation failures. If the failure cannot be solved or explained, alternative models can be applied. For instance, EFMrep, handling replicates independently, or a discrete approach with no model validation [7].

These results are based only on IDseek® OmniSTR™ Global data. As with CE, performance of different kits is likely to vary, and current results might apply differently to other MPS STR kits.

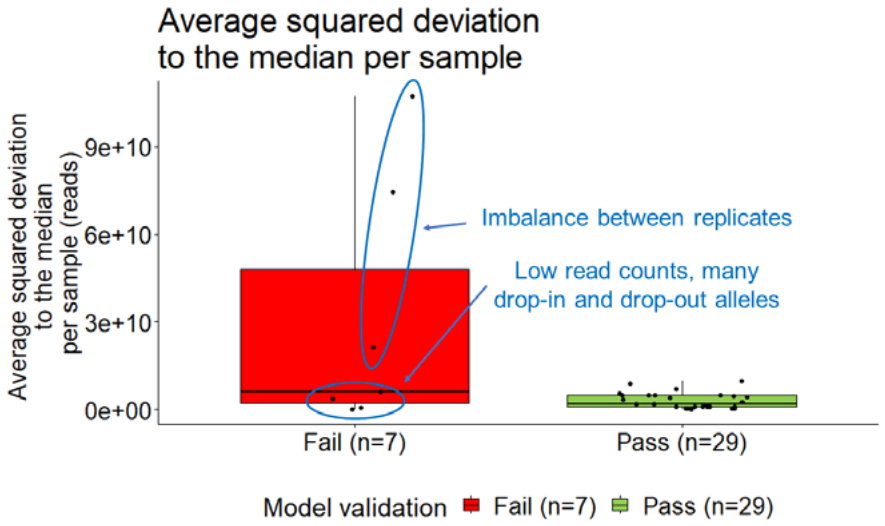


Figure 3. Average squared deviation to the median of the total number of reads between replicates of 36 samples with failed and passed model validation. Hypotheses used: H1 - Major contributor + Minor contributor + Unknown contributor(s), H2 - Major contributor + Unknown contributors.

4. Conclusion

The same CE-based quantitative LR-models can be applied to MPS STR data when considering the assumptions that are made by the models. This study shows that normalization succeeds in meeting these assumptions thereby reducing failed model validation for varying replicates.

5. Acknowledgments

The authors acknowledge Prof. de Knijff from the Leiden University Medical Center for providing DNA sample extracts.

6. Conflict of interest statement

The authors declare no conflict of interest.

7. References

[1] Alonso, A., Barrio, P. A., Müller, P. et al. (2018). Current state>of>art of STR sequencing in forensic genetics. *Electrophoresis*, *39*(21), 2655–2668.

[2] van der Gaag, K. J., de Leeuw, R. H., Hoogenboom, J. et al. (2016). Massively parallel sequencing of short tandem repeats—Population data and mixture analysis results for the PowerSeq™ system. *Forensic Sci Int Genet*. *24*, 86–96.

[3] Benschop, C. C. G., van der Gaag, K. J., de Vreede, J. et al (2021). Application of a probabilistic genotyping software to MPS mixture STR data is supported by similar trends in LRs compared with CE data. *Forensic Sci Int Genet*. *52*, 102489.

[4] Bleka, Ø., Just, R., Agudo, M. M. et al (2022). MPSproto: An extension of EuroForMix to evaluate MPS-STR mixtures. *Forensic Sci Int Genet*. *61*, 102781.

[5] Bleka, Ø., Storvik, G., & Gill, P. (2016). EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. *Forensic Sci Int Genet*. *21*, 35–44.

[6] Benschop, C. C. G., Hoogenboom, J., Hovers, P. et al (2019). DNAxs/DNAStatistX: Development and validation of a software suite for the data management and probabilistic interpretation of DNA profiles. *Forensic Sci Int Genet*. *42*, 81–89.

[7] Bleka, Ø., Prieto, L., & Gill, P. (2022). EFMrep: An extension of EuroForMix for improved combination of STR DNA mixture profiles. *Forensic Sci Int Genet*, *61*, 102771.

[8] Hoogenboom, J., van der Gaag, K. J., de Leeuw, R. et al (2017). FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise. *Forensic Sci Int Genet*. *27*, 27–40.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1013-1019
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Recommendations for identifying disaster Victims from the preliminary results of the "Mass Grave Project": Forensic genetics issues

Giulia Sguazzi[1,2*], Greta Peroni[2], Nicole Colombo[2], Moya Mccarthy-Allen[3], Daniel J. Wescott[4], Timothy P. Gocha[4], Hayley L. Mickleburgh[5], Noemi Procopio[6] and Sarah Gino[2,7]

[1]*CRIMEDIM, Center for Research and Training in Disaster Medicine, Humanitarian Aid and Global Health, University of Eastern Piedmont, Novara, Italy;*

[2]*Department of Health Sciences, University of Eastern Piedmont, Novara, Italy;*

[3]*Division of Biological Traces, Netherlands Forensic Institute, Den Haag, Netherlands;*

[4]*Faculty of Humanities, University of Amsterdam, Amsterdam, Netherlands;*

[5]*Forensic Anthropology Center, Texas State University, San Marcos, TX, USA;*

[6]*Research Centre for Field Archaeology and Forensic Taphonomy, School of Law and Policing, University of Central Lancashire, Preston, UK;*

[7]*Direzione Medica dei Presidi Ospedalieri, AOU Maggiore della Carità, Novara, Italy*

** Corresponding author at: CRIMEDIM, University of Eastern Piedmont, Via Lanino 1, 28100 Novara, Italy, giulia.sguazzi@uniupo.it*

Abstract

Disaster victim identification poses serious challenges to forensic scientists, including bodies or body parts requiring identification in remote places with limited laboratory access. In these settings, samples for genetic analysis may be collected at a later stage, after freezing of the corpses. Alongside tissue sampling and given the ease of collection and the possibility of preservation under harsh conditions, swabs can be aviable and less invasive alternative to bone sampling. To maximize the success of DNA analysis are important sampling as many tissues as possible is recommended, as DNA yield varies by anatomical region.

As part of the Mass Grave Project, an interdisciplinary taphonomic study simulating primary and secondary clandestine mass and single graves with whole body donors at the outdoor research facility of the Forensic Anthropology Center, Texas State University (FACTS), nine body donors were sampled at multiple time points with the aim of proposing sampling strategies and

recommendations for the human identification through DNA analysis. Swabs were collected upon arrival of donors at FACTS (fresh bodies) and after their freezing (between 12-455 days). Subsequently, donors were buried for 18 months, six in a mass grave and three in single graves. After burial, swabs (skin, oral, rectal, periocular) were collected and tissues (muscle, nails, internal organs, cartilage) were sampled from each donor.
The inter/intra-individual variation in DNA quantity and quality was investigated via STR typing. The results showed a progressive loss of information within defrosted and decomposed bodies and were unaffected by the observed degree of decomposition of the tissues nor by the position of the bodies within the mass grave.
Our analyses showed that freezing and burial can affect personal identification, so it is good practice to sample as many biological tissues/fluids as possible, even considering 'uncommon' sampling strategies such as rectal swabs, and to proceed as quickly as possible to successfully identify the remains of unknown individuals when they are discovered.

Keywords
Disaster victim identification, mass fatality, genetics practitioners.

1. Introduction

Mass graves are often grim remnants of conflict and human rights abuses, requiring rapid burial after large-scale losses [1]. Disaster victim identification (DVI) presents significant challenges for forensic scientists, especially when bodies or remains need to be identified in remote locations with limited laboratory access. In such situations, genetic samples might be gathered later, often after the body has been frozen, and frequently from hard tissues, necessitating intricate and perilous extraction procedures [2]. Given these obstacles, recent studies have investigated the feasibility of DNA analysis from soft tissues [3–5]However, identifying the optimal sites for DNA sampling in mass graves remains uncertain [6]. As part of the Experimental Mass Grave Project (MGP), this interdisciplinary taphonomic study simulates primary and secondary clandestine mass and single graves to develop and refine biomolecular methods for human identification in forensic investigations of clandestine graves.

2. Material studied, methods, techniques

This research is part of the MGP conducted at the FACTS. The multidisciplinary team focuses on developing investigative techniques for detection, excavation,

evidence analysis, and documentation in DVI. The May 2021 MGP involved excavating three individual graves and a mass grave containing the bodies of nine donors (six in the mass grave and three buried individually). The research focuses on DNA analysis of biological samples taken from bodies in three different states of decomposition: 'fresh', *at arrival at FACTS (except for individuals D3 and D11, where 'fresh' samples were not collected)*, 'after-freezing' bodies at -20°C *(for a period ranging from 12 to 455 days for the various donors)*, and 'post-burial', after 18 months. Ethical approval and permission were obtained from FACTS to collect human tissue in accordance with the Texas Uniform Anatomical Gift Act (Health and Safety Code Chapter 692A); and for DNA analysis from the samples collected in Italy by the Novara Intercompany Ethics Committee (CE 24/21) and in the United Kingdom by the Northumbria University Ethics Committee (submission ref. 24514 and 29218).

2.1. Sample collection

The experimental involves the DNA analysis of 187 samples collected at three time points: 35 'fresh' swabs, 45 'after-freezing' swabs and 107 'post-burial' samples. At each sampling time, swab samples were collected from various body locations including oral, rectal, neck, hand and foot. After burial, regardless of the body's condition, additional biological material was collected. In particular periocular swabs were performed, as well as sampling of different tissues such as skin, muscle, internal organs, cartilage, and nails.

2.2. Sample analysis

DNA was extracted from samples using the QIAamp DNA Mini Kit and EZ2 Connect and EZ1&2 Tissue Kit (QIAGEN) for periocular swab and tissue, optimizing the protocol recommended by the manufacturers. DNA quantification was performed using the Quantifiler Trio DNA Quantification (ThermoFisher Scientific) on a 7500 Fast Real-Time PCR System (Applied Biosystems). A total of 131 extracts containing at least one picogram of human DNA were amplified using the GlobalFiler PCR Amplification Kit (ThermoFisher Scientific) and sequenced using the SeqStudio Genetic Analyzer (ThermoFisher Scientific). The resulting STR profiles were analysed using GeneMapper® ID-X v1.5 software. To evaluate the goodness of the obtained profiles, we applied the criterion distinguishing between profiles with fewer than 7 loci and those with more than 10 loci. This criterion aligns with the guidelines specified in DPR 87 of 7 April 2016 for inclusion in the National DNA Database and comparison of genetic profiles in Italy.

3. Results and Discussion

Following DNA extraction, amplification, and sequencing, genetic profiles were typed for 73% of samples. The success rates, measured by the number of loci successfully typed, were compared across the three experimental conditions. The results, summarised in succession rate when the number of loci greater than 10 and failures if the loci are less than 7, are presented in **Figure 1**.

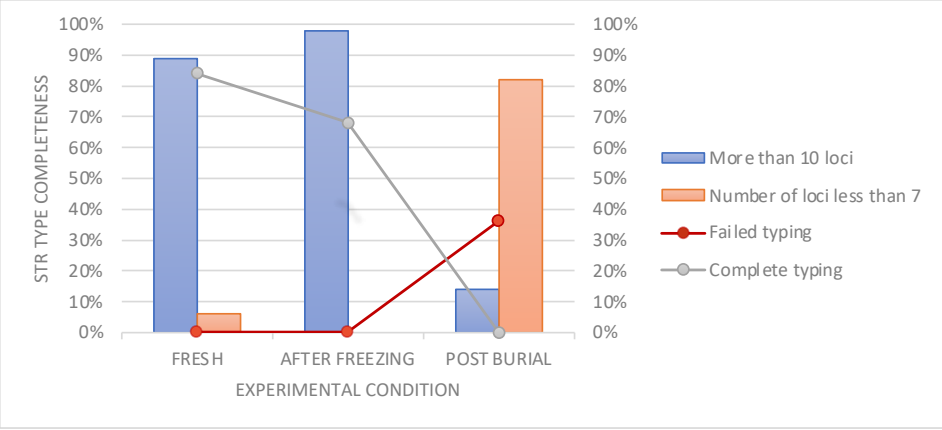
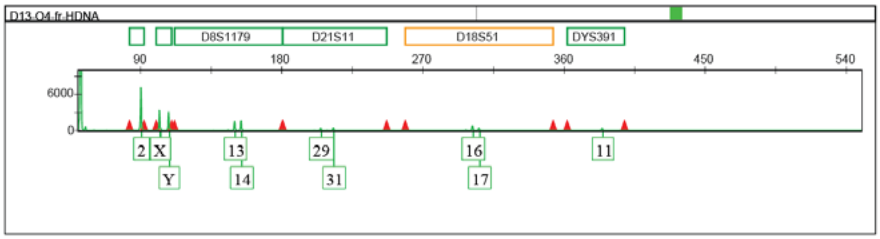


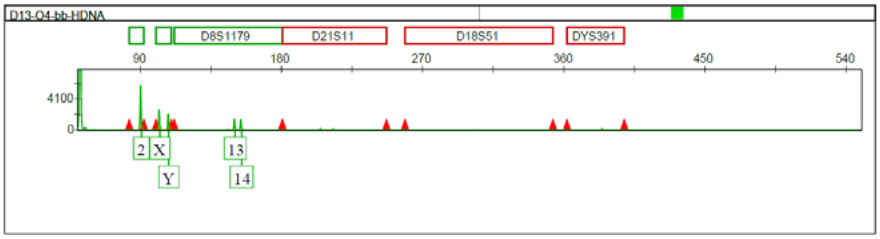
Figure 1. Histogram of success rates (loci >= 10) and failure rates (loci < 7) of STR typing in the three experimental conditions.

Amongst the 131 amplified extracts, two provided noteworthy results. The sample D13_R6_de_HDNA, a rectal swab taken 18 months after burial, resulted in a complete genetic profile, whereas the 'post burial' buccal swab did not provide a sufficient amount of DNA (**Figure 2**).

A



B



C

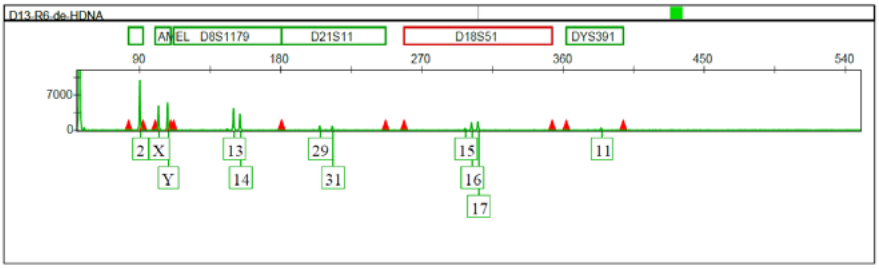


Figure 2. Comparison of the electrophorogram (only the green channel is shown) of two oral swabs (A and B) and one rectal swab (C) taken from donor D13 under the three conditions (A 'fresh', B 'after freezing' and C 'post burial').

The sample D23_F4_fr_HDNA, a skin swab taken from the foot's sole of the female donor D23 on arrival at FACTS, (of which the green channel is shown in Figure 3) resulted in a mixed profile with a female minority component from the D23 donor and an unattributed majority male component. The male component profile was compared with other male donors in the study and with the researchers involved in laboratory analysis.

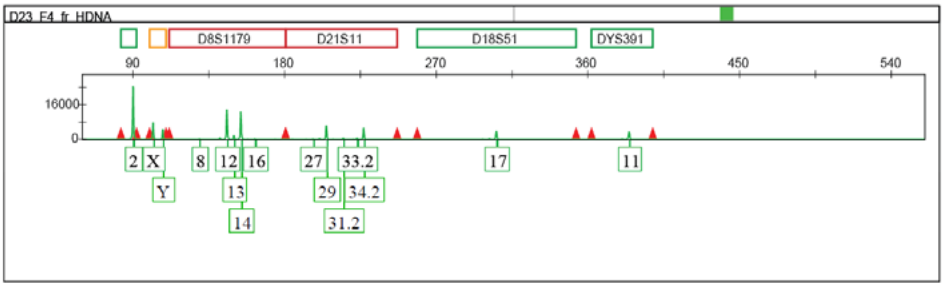


Figure 3. Electrophorogram of the sample D23_F4_fr_HDNA for the green channel.

4. Discussion

Our preliminary results suggest that freezing and burial deceased bodies can lead to a gradual loss of genetic information. Additionally, to overcome the challenges associated with decomposition and contamination, sampling a wide range of tissues, including unusual samples like rectal swabs, is crucial. This approach complements traditional bone and tooth analysis. Moreover, the example of the mixed profile obtained from donor D23 illustrates that it is not only the catastrophic event itself that can be a source of contamination, but that sampling and handling in the autopsy room can equally introduce genetic contaminations.

5. Conclusion

Freezing and burial in challenging contexts can hinder the identification of remains. To overcome these challenges, swab samples can be a valuable alternative to tissue and bone samples due to their ease of collection and preservation. *As DNA yield varies by anatomical region, it is advisable to* prioritize protected areas to minimize contamination. *Collecting as much biological material types as possible, and doing so promptly, is crucial for identifying unknown bodies. Appropriate training for* disaster scene personnel*, including knowledge of forensic practices, is essential to ensure that evidence is not compromised.*

6. Acknowledgments

The authors express their gratitude to the donors and their families for allowing the use of donated bodies in this research, as well as the Forensic Anthropology Research Facility (FARF).

This article is the result of a study conducted in the framework of the International PhD in Global Health, Humanitarian Aid, and Disaster Medicine organized by University of Eastern Piedmont (UPO).

7. Conflict of interest statement

The authors declare no conflicts of interest.

8. References

[1] S.I. Ashbridge, P.S. Randolph-Quinney, R.C. Janaway, S.L. Forbes, O. Ivshina, Environmental conditions and bodily decomposition: Implications for long term management of war fatalities and the identification of the dead during the ongoing Ukrainian conflict, Forensic Sci. Int. Synerg. 5 (2022). https://doi.org/10.1016/j.fsisyn.2022.100284.

[2] L.G. Biesecker, J.E. Bailey-Wilson, J. Ballantyne, H. Baum, F.R. Bieber, C. Brenner, B. Budowle, J.M. Butler, G. Carmody, P.M. Conneally, B. Duceman, A. Eisenberg, L. Forman, K.K. Kidd, B. Leclair, S. Niezgoda, T.J. Parsons, E. Pugh, R. Shaler, S.T. Sherry, A. Sozer, A. Walsh, DNA identifications after the 9/11 World Trade Center attack, in: Science (80-.)., 2005: pp. 1122–1123. https://doi.org/10.1126/science.1116608.

[3] V. Agostini, S. Gino, S. Inturri, A. Marino, N. Staiti, M. Sticchi, E. Chiti, P. Linarello, G. Gentile, P. Primignani, M. Giriodi, P. Bailo, A. Piccinini, "UNUSUAL" TISSUES AND SAMPLE COLLECTION STRATEGIES ON EXHUMED BODIES, Forensic Sci. Int. Genet. Suppl. Ser. 7 (2019) 169–171. https://doi.org/10.1016/j.fsigss.2019.09.066.

[4] M. Allouche, M. Hamdoum, P. Mangin, V. Castella, Genetic identification of decomposed cadavers using nails as DNA source, Forensic Sci. Int. Genet. 3 (2008) 46–49. https://doi.org/10.1016/j.fsigen.2008.07.008.

[5] A. Trindade-Filho, C. Mendes, S. Ferreira, S. Oliveira, A. Vasconcelos, F. Maia, H. Pak, K. Paula, DNA obtained from decomposed corpses cartilage: A comparison with skeleton muscle source, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2008) 459–461. https://doi.org/10.1016/J.FSIGSS.2007.10.210.

[6] A. Senst, E. Scheurer, K. Gerlach, I. Schulz, Which tissue to take? A retrospective study of the identification success of altered human remains, J. Forensic Leg. Med. 84 (2021) 102271. https://doi.org/10.1016/j.jflm.2021.102271.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1021-1025
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Development of a Nicking Endonuclease Dependent Amplification Method and Its Use in STR Analysis of Trace DNA

Wanli Bi[1]

[1] *NUHIGH Biotechnologies Co., Ltd. Suzhou, China.*

Abstract
The detection of trace DNA from crime scenes is a major challenge in forensic DNA analysis, often resulting in allele dropout or complete amplification failure. To address these issues, we developed a primer-free isothermal whole-genome amplification method called Nicking Endonuclease Dependent Amplification (NDA). NDA maintains high fidelity and preserves the original genotype of trace DNA without introducing errors. Two STR amplification systems—17-loci autosomal STR and 23-loci Y-STR—were optimized for use with NDA. Sensitivity tests revealed detection limits of 3 pg/μL for autosomal STR and 4 pg/μL for Y-STR. Amplification efficiency improved by 2- to 33-fold across loci. NDA also demonstrated robust tolerance to common PCR inhibitors, including humic acid, hematin, and collagen, establishing it as a reliable method for trace DNA analysis in forensic science.

Keywords
Nicking endonuclease, Pre-amplification, STR, Trace DNA.

1. Introduction

Short tandem repeat (STR) typing is crucial in forensic DNA analysis, particularly for individual identification and paternity testing. However, the analysis of trace DNA from crime scenes is often hindered by low DNA quantities, leading to amplification failures or incomplete profiles. Conventional pre-amplification methods, including whole genome amplification (WGA), multiple displacement amplification (MDA), and rolling circle amplification (RCA), have been used to enhance trace DNA detection. These methods, however, are prone to errors due to primer use, resulting in amplification artifacts and biases that compromise STR profiling.

To address these challenges, we developed Nicking Endonuclease Dependent Amplification (NDA), a primer-free isothermal method ensuring linear amplification with minimal error rates. NDA employs the nicking endonuclease Nt.BstNBI and *Bst* DNA polymerase to amplify DNA through repeated strand nicking and extension, eliminating primer-dimers and non-specific amplification, making it highly suitable for forensic applications (Fig. 1).

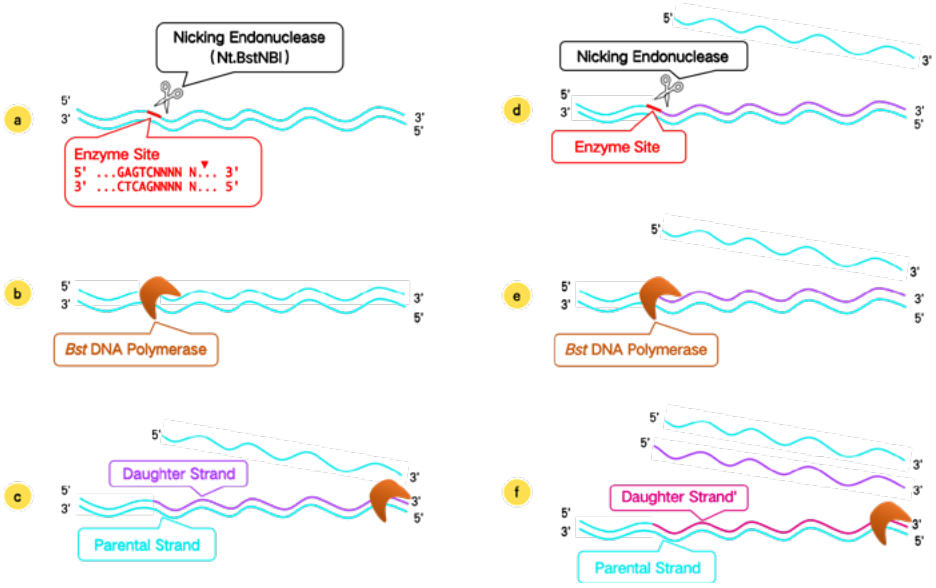


Figure 1. The working principle of NDA

2. Material and methods

2.1. STR Primer Design

Sequence analysis identified 17 autosomal STRs and 23 Y-STRs with flanking regions compatible with NDA. These loci were grouped into two panels: TraceA (17 autosomal STRs and Amelogenin) and TraceY (23 Y-STRs). The primers were designed for optimal NDA compatibility.

2.2. Sensitivity Tests

Standard 9948 DNA was prepared at 1, 2, 3, 4, and 5 pg/μL concentrations. A 20 μL NDA reaction was conducted at 59°C for 40 minutes, followed by enzyme

inactivation at 95°C for 5 minutes. PCR amplification of NDA products was performed under standard thermal cycling conditions.

2.3. Amplification efficiency

To assess amplification efficiency, 0.1 ng of 9948 DNA was amplified using NDA. Peak heights of STR alleles were compared between NDA-treated and untreated samples to calculate efficiency improvements.

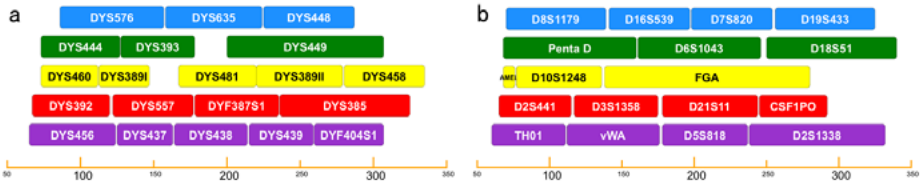
2.4. Robustness against inhibition

NDA's tolerance to common PCR inhibitors was tested by adding humic acid, hematin, and collagen to the reaction at varying concentrations. Amplification performance was assessed based on allele detection rates and peak heights post-PCR.

3. Results and discussion

3.1 STR Primer Design

Sequence analysis identified 17 autosomal STRs and 23 Y-STRs compatible with NDA amplification. These loci were successfully incorporated into two panels—TraceA and TraceY—enabling linear amplification of trace DNA (Fig. 2).



Figre 2. Panel design of auto-STRs and Y-STRs matched with NDA (a. auto-STRs, b. Y-STRs)

3.2 Sensitivity

NDA significantly improved STR amplification sensitivity. TraceA achieved full allele detection at 3 pg/μL, while TraceY detected 97.8% of alleles at 3 pg/μL and 100% at 4 pg/μL. Conventional PCR required at least 10 pg/μL for similar profiles.

Table 1. Sensitivity Results of Trace A and Trace Y with NDA

DNA Concentration (pg/μL)	TraceA		TraceY	
	Allele Detection Rate	Peak Height (RFU)	Allele Detection Rate	Peak Height (RFU)
1	90.3%	2 522.19	78.3%	1 121.24
2	95.2%	2 710.08	84.8%	2 211.08
3	100%	3 517.47	97.8%	2 498.22
4	100%	3 807.72	100%	2 820.82
5	100%	4 531.08	100%	3 423.17

3.3 Amplification efficiency

NDA increased amplification efficiency by an average of 5-fold for autosomal STRs and 7-fold for Y-STRs, with improvements ranging from 2.03 to 24.75 for autosomal loci and 2.04 to 33.95 for Y-STR loci (Fig. 3).

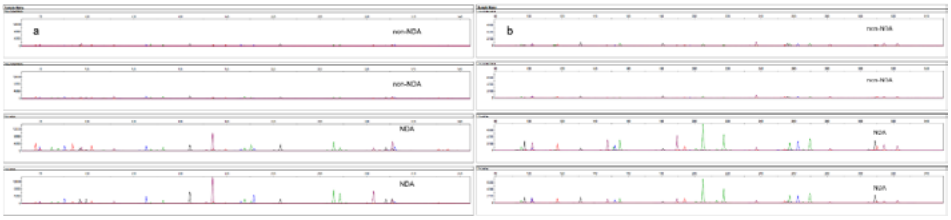


Figure 3. Comparison of peak height between NDA and non-NDA (a. auto-STRs, b. Y-STRs)

3.4 Robustness against inhibition

NDA demonstrated superior robustness against PCR inhibitors. Even at high concentrations of humic acid, hematin, and collagen, NDA maintained high allele detection rates and peak heights, confirming its suitability for challenging forensic samples (Fig. 4).

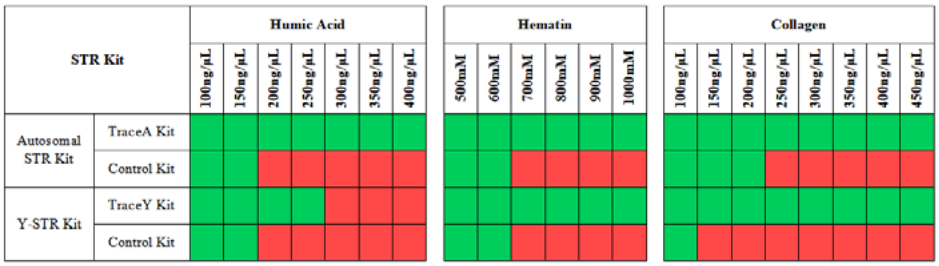


Figure 4. Results of Trace A and Trace Y with NDA against inhibition.

4. Conclusion

NDA is a reliable, high-fidelity pre-amplification method that enhances trace DNA analysis. By eliminating primer-induced errors and providing strong tolerance to inhibitors, NDA ensures accurate STR profiling even with highly degraded or inhibited samples. Its successful integration with 17 autosomal STRs and 23 Y-STRs highlights its potential as a valuable tool for forensic applications.

5. Acknowledgments

We would like to thank Nantong Public Security Berean for supporting this research.

6. Conflict of interest statement

The authors declare no conflicts of interest.

7. References

[1] L. Deleye, P.A.S. Vander, J. Weymaere, et al., Short tandem repeat analysis after whole genome amplification of single B-lymphoblastoid cells, Sci. Rep. 8 (2018) 1-6.

[2] S. Cao, X. Tang, T. Chen, et al., Types and applications of nicking enzyme-combined isothermal amplification, IntJ. Mol. Sci. 23 (2022) 4620.

[3] L. Wang, C. Qian, H. Wu, et al., Technical aspects of nicking enzyme assisted amplification, Analyst, 143 (2018) 1444-1453.

[4] C. Qian, R. Wang, H. Wu, et al., Nicking enzyme-assisted amplification (NEAA) technology and its applications: a review, Anal. Chim. Acta. 1050 (2019) 1-15.

[5] S.H. Chan, B.L. Stoddard, S.Y. Xu. Natural and engineered nicking endonucleases—From cleavage mechanism to engineering of strand-specificity, Nucleic. Acids. Res. 39 (2011) 1-18.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1027-1033
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Optimizing the Recovery of DNA from Exhaled Breath Devices for Human Identification

Kayli Carrillo[1], Sheree Hughes[1*]

[1] *Department of Forensic Science, Sam Houston State University, Huntsville, TX 77340, USA*

** Corresponding author at: Department of Forensic Science, Sam Houston State University, 1003 Bowers Blvd., Huntsville, TX 77340, USA, sxh039@shsu.edu*

Abstract

Recovering DNA from exhaled breath offers potential forensic applications but presents distinct challenges. This study aimed to assess whether DNA could be captured from exhaled breath using two breath collection devices: Breath Explor® and SensAbuse®. These devices are typically employed for in-field drug testing, with samples sent by mail to laboratories for analysis, emphasizing the importance of maintaining sample integrity to safeguard the chain of custody.

Various DNA collection and extraction methods (swabbing the mouthpieces with cotton and microFLOQ™ swabs, swabbing the filters with microFLOQ™ swabs or soaking filters with DNA Investigator® kit chemistry or following the San Diego Police Department protocol) were evaluated to determine if quality STR profiles could be obtained from these devices' mouthpieces and/or internal filters. All samples were quantified using the Quantiplex Pro kit and genotyped with the Investigator® 24 plex kit on a 3500 Genetic Analyzer. However, these approaches yielded suboptimal DNA recovery and partial STR profiles.

In an effort to improve the capture of DNA in exhaled breath, wet or dry FTA® card punches were inserted into the Breath Explor® device's mouthpiece. This strategy failed to yield any significant improvements, with less than 10% of samples exhibiting detectable DNA quantities. Additionally, the use of Diamond™ Dye was investigated in an attempt to visualize the location of cell or cell-free DNA transfer from exhaled breath onto the devices, allowing for targeted sample collection.

In summary, this study confirms that collecting DNA from exhaled breath is challenging; however, in cases where the person being drug tested with these devices is in question, laboratories should prioritize swabbing the inside of the mouthpiece for the most successful and reliable STR typing results.

Keywords

Low-template DNA, Exhaled Breath DNA

1. Introduction

Exhaled breath, though primarily used for diagnosing lung disease and detecting volatile illicit drug use, remains a relatively unexplored DNA matrix. Its composition consists of mediators and nucleic acids, which are released by apoptosis, necrosis, and spontaneous cell death in the respiratory tract due to oxidative stresses [1].

Several breath devices used for drug detection include Breath Explor® and SensAbues®. Breath Explor® collects aerosol products of the surfactant from the distal areas of the lungs through impaction of the three internal filters [2]. SensAbues® contains a thin electret polymer air filter that captures and retains bioaerosol particles from the airway lining fluid of the lungs [3].

This study aims to assess different DNA collection methods, including the use of different swabs, various soaking techniques, incorporation of FTA® card punches, and cell visualization using Diamond™ Dye, to determine whether usable STR profiles can be obtained from the mouthpiece and/or internal filters of the devices to confirm sample integrity.

2. Material studied, methods, techniques

2.1. Sample Collection

Exhaled breath samples from ten donors were collected following SHSU IRB-approved procedures (IRB-2021-275). Buccal swabs served as DNA reference samples, and all were de-identified with unique numbers upon receipt.

According to the manufacturer's guidelines for SensAbues® sample collection, participants were required to inhale and exhale twenty times. For Breath Explor®, the participants were instructed, per the manufacturer's recommendations, to first exhale completely to empty the lungs, hold their breath for three seconds, and then take a deep breath through the device for eight seconds. This process was repeated twelve times.

2.2. DNA collection and extraction methods from mouthpieces and filters

Five different collection and extraction methods were evaluated for the DNA recovery of the mouthpieces and filters of both devices. Ten participants provided breath samples for both device types for each method (N = 100). Mouthpieces were either swabbed with a cotton swab moistened with sterile water followed by extraction using the EZ1 DNA Investigator® Kit (QIAGEN, Germantown, MD) and were

purified with the EZ1 Advanced XL (QIAGEN) using the large volume protocol with an elution volume of 40µl. The other swabbing method consisted of using a microFLOQ™ (COPAN Diagnostics Inc. Murrieta, CA) swab and was processed with direct PCR following the Investigator® 24plex GO! Kit (QIAGEN) protocol. For the filters, samples were either swabbed with a microFLOQ™ swab for direct PCR analysis or subjected to two different soaking methods: one using EZ1 DNA Investigator® kit chemistry and the other following the San Diego Police Department (SDPD) protocol [4].

To prepare for all soaking methods, SensAbues® filters were cut in half and then cut into smaller pieces to fit inside three separate 2mL tubes. SensAbues® filters were soaked following the previously published SDPD method [4] and purified on the EZ1 Advanced XL following the large-volume protocol. The three elution tubes were combined into one 1.5 mL tube and were concentrated using the Savant SpeedVac DNA 130 Vacuum Concentrator (ThermoFisher Scientific, Waltham, MA) at 35°C for 90 minutes. Samples were then reconstituted with 40µL of water.

The Breath Explor® device has three separate filters that were individually treated before combining into one sample for all soaking methods. The filters were also soaked following the previously described SDPD protocol and purification methodology. The three elution tubes were similarly combined into one tube, concentrated with the same parameters, and reconstituted in 40µL.

Filters for both SensAbues® and Breath Explor® were soaked using EZ1 DNA Investigator® chemistry. Filters were soaked with 475µL of G2 Buffer (QIAGEN) and 25µL of proteinase K and were incubated at 56°C for 1 hour. Before purification, 400µL of MLT buffer and 1µL of carrier RNA were added to the lysate. Filters were purified on the EZ1 Advanced XL platform following the large volume protocol. Samples for both filter types were concentrated and reconstituted using the same protocol as mentioned previously.

2.3. DNA capture from Exhaled breath samples using FTA® card punches

Wet and dry FTA® card punches were tested for their potential to capture DNA from exhaled breath samples. Hole punches (19mm) were taken from QIAcard FTA® micro cards and were placed inside the Breath Explor® device. Three volunteers provided samples for five devices – half of the devices contained pre-wetted FTA® card punches, while the other half contained dry punches (n = 30). FTA® cards were left to dry overnight and were lysed in 290µl of G2 buffer and 10µL of proteinase K at 56°C at 900 rpm for 90 minutes. Samples were purified on the EZ1 Advanced XL following the tip dance protocol.

2.4. DNA quantification and amplification

DNA extracts were quantified using the Quantiplex® Pro Kit (QIAGEN) per manufacturer recommendations on the 7500 real-time PCR instrument (ThermoFisher Scientific). DNA extracts were amplified using the Investigator® 24plex PCR Amplification Kit (QIAGEN) following the manufacturer's protocol on the ProFlex PCR System (ThermoFisher Scientific).

2.5. Separation, detection, and analysis

Amplified products were separated and detected on the ABI 3500 Genetic Analyzer (ThermoFisher Scientific). GeneMapper® ID-X v1.4 software (ThermoFisher Scientific) was used to analyze the data with an analytical threshold of 100 RFUs and a stochastic threshold of 200 RFUs. Completeness of profiles was assessed by comparing the number of reportable alleles to the reference profile.

2.6. Evaluation of Diamond Dye for use of visualizing cells for breath devices

Diamond™ Dye (Promega, Madison, WI, USA) was tested for its ability to visualize DNA. As a positive control, 5µL of saliva was spiked on a glass slide, while 5µL of water was used as a negative control on a separate slide. A solution of 20X Diamond™ Dye and 75% ethanol was sprayed onto the substrate to determine if cells could be visualized. Substrates were examined with an excitation wavelength of 494 nm using the Nightsea Blue light excitation lamp and an emission wavelength of 555 nm using a translucent filter visualized with a Keyence VHX-7000 microscope.

3. Results and discussion

Both filter types yielded picogram or sub-picogram amounts of DNA. The highest-yielding sample was from a SensAbues® device using the SDPD soaking method with 0.26 ng. No statistical difference was observed between the two soaking methods for Breath Explor ($p = 0.2$) and SensAbues ($p = 0.68$). The observed average percentage of reportable alleles was less than 20% from the filters of both devices (Fig 1). No STR profile was obtained for 76% of filter samples, and only 6% of filters yielded a complete profile ($n = 60$) (data not shown).

As expected, swabbing from the mouthpiece of the devices yielded more DNA than from the internally located filters (Fig 1). Overall, an average of 99.35% of alleles were generated from all swabs from the mouthpiece of both devices. Complete STR profiles were generated from all cotton swabs used to sample the Breath

Explor® devices and all microFLOQ® swabs from the SensAbues® devices. Near complete DNA profiles were recovered from microFLOQ® swabs of Breath Explor® devices and cotton swabs of SensAbues® (average 99.9% and 97.5% allele recovery, respectively).

Due to the insufficient amounts of DNA recovered from the internal filters, these results prompted the investigation of an alternative approach for collecting DNA from exhaled breath by modifying the Breath Explor® device by adding an FTA® card punch directly inside the mouthpiece. Only Breath Explor® was tested due to the ease with which the punch could be added to the filter. While FTA® cards are traditionally used to store blood or buccal samples, several studies have explored their use for collecting trace DNA [5]. However, no DNA was detected in any of the dry FTA® cards, while only one out of five of the wetted FTA cards from each volunteer yielded detectable amounts of DNA (data not shown). No PCR inhibition was observed in any of the samples, but due to the low amounts of total DNA, this method did not improve DNA collection from breath devices and was not considered suitable for downstream STR analysis and further investigations. The pore size for FTA® cards ranges from 0.1μM to 0.9 μM [6], while aerosol particles range from a size of 0.3μM to 2.0μM [7], meaning that only a portion of particles could be captured by the filter thus giving a possible explanation for the unsuccessful capture of DNA. Therefore, investigation of other filters may show more success as shown in a study demonstrating the recovery of human DNA in air in rooms [8].

Lastly, the visualization of cells using Diamond™ Dye with the positive control using saliva deposited on a glass slide did not yield successful results, as no clear distinction could be made between the positive and negative controls (data not shown). Difficulties of using Diamond Dye® included autofluorescence and the lack of contrast between the breath device substrates as Breath Explor® is a green device and SensAbues® is a blue device. Further investigation using a Polilight lamp and Dino-Lite fluorescence digital microscope (AnMo Electronics Corporation, New Taipei City, TWN) will be examined as several previous studies using Diamond™ Dye and the Dino-Lite microscope have been reported for the use of visualizing trace DNA [9,10].

4. Conclusion

This research highlights the challenges of recovering DNA from exhaled breath. Retrieving DNA from the filters of both Breath Explor® and SensAbues® proved

unreliable, and incorporating FTA® card punches did not improve DNA capture. Further studies are needed to evaluate the use of Diamond™ Dye for visualizing trace DNA on breath devices. It is recommended that laboratories use traditional methods, such as swabbing the mouthpiece and routine DNA extraction, as swabbing provides a simpler workflow compared to soaking filters. Ultimately, we demonstrated an effective method for maintaining the chain of custody when needed.

5. Acknowledgments

The authors would like to acknowledge the Department of Forensic Sciences, the volunteers, and SensAbues® for providing the devices free of charge.

6. Conflict of interest statement

There are no conflicts of interest.

7. References

[1] O. Beck, K. Leine, G. Palmskog, et al. Amphetamines detected in exhaled breath from drug additcs: A new possible method for drugs-of-abuse testing, J. Anal. Tox. 34 (2010) 233-7.

[2] V. Malave, K. Jeerage, E. Garboczi, et al. 3D computational fluid and particle dynamics simulations: metrics of aerosol capture by impaction filters. J. Breath Res. 18 (2023) 016002.

[3] S. Sefaraj, S. Ullah, A. Tinglev, et al. Evaluation of a new simple collection device for sampling of microparticles in exhaled breath, J. Breath Res. 12 (2018) 036005.

[4] S. Montepetit, P. O'Donnell, An optimized procedure for obtaining DNA from fired and unfired ammunition, Forensic Sci. Int. Genet. 17 (2015) 70-74.

[5] I. Kirgiz, C. Calloway, Increased recovery of touch DNA evidence using FTA paper compared to conventional collection methods, J. Forensic and Leg. Med. 47 (2017) 9-15.

[6] P. Kanokwongnuwut, P. Kirkbride, A. Linacre, Latent DNA detection, Forensic Sci. Int. Genet. 37 (2018) 95-101.

[7] J. Champion, P. Kanokwongnuwut, R.van Oorschot, Evaluation of a fluorescent dye to visualise touch DNA on various substrates, J. Sci. 66 (2021) 1435-1442.

[8] M. Goray, D. Taylor, E. Bibbo, et al. Up in the air: Presence and collection of DNA from air and air conditioner units, Electrophoresis. 45 (2023) 933-947.

[9] R. Tang, M. Li, X. Yan, et al. Comparison of paper-based nucleic acid extraction materials for point-of-care testing applications, Cellulose (Lond). 29 (2022) 2479-2495.

[10] A. Almstrand, E. Ljungstrom, J. Lausmaa, et al. Airway monitoring by collection and mass spectrometric analysis of exhaled particles, Anal. Chem. 81 (2009) 662-668.

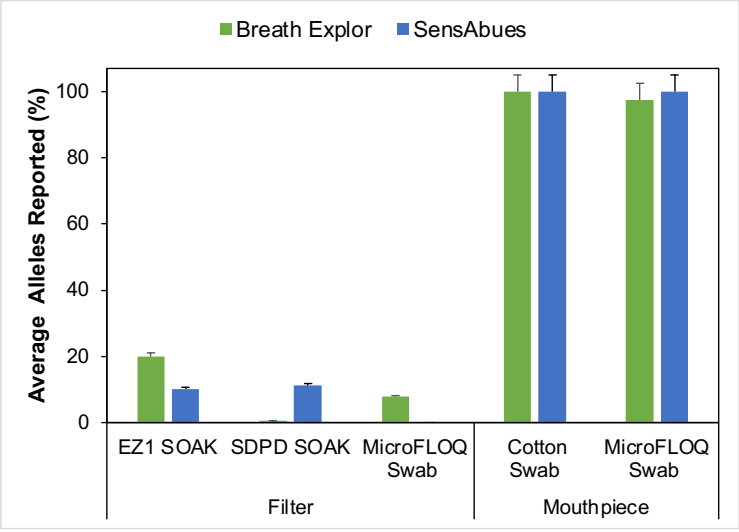


Figure 1. Average reportable alleles (%) versus each extraction and collection method for Breath Explor® and SensAbues® (n = 100). Error bars represent percent error (5%).

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1035-1042
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Comparing Extraction and Direct Amplification Methods for Enhanced Touch DNA Profiling

Salem K. Alketbi[1,2,3*], Will Goodwin[3], Hussain J. Alghanim[1], Sheikha H. Sanqoor[1], Suaad A. Alshehhi[1], Maryam M. Almheiri[1], Naeema S. AlJanaahi[1], Afra N. Sanqoor[1], Fatma J. Altamimi[1], Mariam M. Sajwani[1]

[1]*The Biology and DNA Section, General Department of Forensic Science and Criminology, Dubai Police General Head Quarters, Dubai, UAE.*

[2]*International Center for Forensic Sciences, Dubai Police General Head Quarters, Dubai, UAE.*

[3]*University of Central Lancashire, Preston, UK.*

** Corresponding author Tel: 00447774141205; E-mail: alkitbe.11@hotmail.com*

Abstract

This study compares traditional DNA extraction and direct amplification methods for Touch DNA profiling across various surface sizes and materials. DNA yield and profile quality were evaluated using RFU signal strength and allele detection. Results show that direct amplification—using MicroFLOQ™ Direct swabs and SceneSafe Fast™ minitape—performs better on smaller surfaces ($p < 0.05$), enabling faster DNA recovery. In contrast, extraction is superior for larger surface areas ($p < 0.001$), highlighting the complementary roles of these methods based on forensic context. These findings provide guidance for optimizing Touch DNA profiling to enhance efficiency and accuracy in forensic investigations.

Keywords

Forensic Genetics, Touch DNA, DNA extraction, Direct amplification, MicroFLOQ™ Direct swabs, SceneSafe Fast™ Minitape.

1. Introduction

Trace DNA profiling plays a vital role in linking individuals to criminal activities. However, Touch DNA presents unique challenges that distinguish it from other forms of DNA evidence [1-3]. Touch DNA, which typically consists of small quantities

of cellular material left behind by contact, is influenced by various factors, including surface types and environmental conditions, complicating its recovery and analysis [4-5]. Unlike body fluids, which generally provide sufficient DNA for analysis, Touch DNA samples often yield limited quantities, making them more difficult to work with [2].

Traditional DNA extraction methods, particularly those involving column-based purification, can exacerbate this challenge by causing further loss of DNA, jeopardizing the profiling of low-copy-number or degraded samples [1,6]. In response to these limitations, there is growing interest in direct amplification protocols that bypass the extraction and quantification steps, streamlining the DNA profiling process for trace samples [7-8].

One promising solution is the microFLOQ® Direct swab, co-developed by the French Gendarmerie Forensic Research Institute (IRCGN™) and Copan [8]. This tool has shown efficacy in collecting trace DNA for direct amplification, reducing DNA loss during the profiling process [8-9]. Additionally, the SceneSafe Fast™ minitape (K545) has proven effective in collecting trace DNA from fabric, with the added advantage of facilitating the simultaneous recovery of DNA and fingerprints [10-14].

This study aimed to explore the performance of two approaches—direct amplification versus traditional extraction—across different surface areas and deposition types. By comparing RFU signal strength and analyzing the resulting DNA profiles, this work provided insights into the most effective methods for DNA profiling from Touch DNA samples in various forensic scenarios.

2. Materials and methods

2.1. Experimental setup and deposition

The deposition procedure followed established protocols [4]. Two participants, categorized as high and low DNA shedders, began by thoroughly washing their hands with antibacterial soap and resting for 10 minutes. Afterward, they induced eccrine sweat by gently touching their foreheads to collect epithelial cells [2]. Each participant then applied moderate pressure using the index, middle, and ring fingers of both hands to deposit DNA onto surfaces of two different sizes: 2.5 × 3.5 cm and 5 × 7 cm.

To ensure the absence of cross-contamination, each deposition was conducted separately for each participant. Touch DNA was deposited on a variety of surfaces, including stainless steel, textured plastic, textured wood, copier paper, fabric (65% polyester, 35% cotton), and glass. Non-porous surfaces were sterilized

using a 2% Virkon solution (a viricidal disinfectant) and exposed to ultraviolet (UV) radiation for 20 minutes. Porous surfaces underwent 30 minutes of UV exposure. All procedures were carried out at room temperature under controlled conditions.

2.2. DNA recovery

Deposited DNA was promptly collected utilizing Copan nylon flocked swabs (4N6FLOQSwabs®) (NS) and SceneSafe Fast™ minitape (K545) (MT) for subsequent extraction. MicroFLOQ™ Direct swabs (MF) were also employed in conjunction with MT for direct amplification. The NS swabs were moistened with 30 µL of sterile distilled water, as per the manufacturer's instructions, while the MF swabs were prepared by adding 1 µL of sterile distilled water using a pipette. The MT did not require additional preparation. To optimize DNA collection, each minitape was applied to the deposition area 16 times [9]. A total of 144 samples were collected, with 72 samples designated for extraction and 72 for direct amplification. The sampling process included six replicates for each of the two deposition area sizes across the six surfaces selected for the study.

2.3. DNA profiling and analysis

Samples collected with Copan nylon flocked swabs (NS) and SceneSafe Fast™ minitape (MT) were extracted using the PrepFiler Express BTA™ kit with the AutoMate Express, using 460 µL of lysis buffer instead of 230 µL, as per the manufacturer's instructions [2]. For NS, complete swab heads were placed into tubes, while for MT, the lower adhesive portion was used, with a final elution volume of 50 µL. Quantification was performed using the Quantifiler® Trio DNA Quantification Kit and QuantStudio 5 Real-Time PCR system, following manufacturer protocols.

For direct amplification, MF swabs were bent directly into PCR tubes (0.2 ml), and triangular portions of MT (~2 mm^2) were inserted. PCR tubes were prepared with 10 µL of PCR master mix and 15 µL of TE buffer.

DNA amplification was conducted using the GlobalFiler™ PCR Amplification Kit on an ABI GeneAmp® 9700 PCR System for 29 cycles. Amplified products were separated and detected using an ABI 3500 Genetic Analyzer with standard injection settings (1.2 kV, 24 s). Analysis was performed with GeneMapper® ID-X Software v1.5, with a detection threshold of 75 RFUs.

Electrophoresis was carried out on a 36-cm capillary array using POP-4™ polymer. Controls confirmed no contamination. Statistical analysis (ANOVA) was conducted in RStudio. RFUs for homozygous loci were reported directly, while heterozygous loci peak heights were summed. All samples yielded complete, single-source DNA profiles.

3. Results

Significant differences in the average signal strength (RFU) were observed when comparing direct amplification methods (MT and MF) with traditional DNA extraction, particularly for samples collected from smaller surface areas (2.5 × 3.5 cm) ($p < 0.05$). Direct amplification using MT and MF yielded higher peak heights compared to the extracted samples (RFU mean: MT = 5779, MF = 7354, EXT = 3703) (see Figure 1). This indicates that MT and MF are more effective in collecting Touch DNA from smaller surface areas, where the concentration of cellular material is higher.

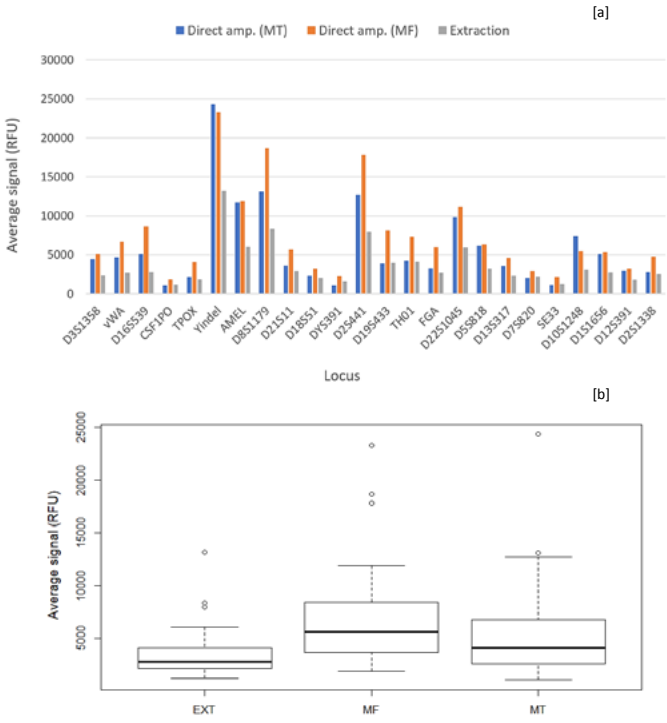


Figure 1. Comparison of the average signal (RFU) per locus for samples [a] and the mean signal (RFU) per DNA profile for samples [b] (n=144) collected from a small surface area **(2.5 × 3.5 cm).** The figure contrasts samples processed via direct amplification using MicroFLOQ™ Direct swabs (MF) and SceneSafe Fast™ minitape (MT), both amplified with the GlobalFiler™ kit, against samples subjected to extraction (EXT) using the PrepFiler Express BTA™ kit, followed by amplification with the GlobalFiler™ kit.

However, when samples were collected from larger surface areas (5 × 7 cm), extracted samples displayed significantly higher peak heights compared to those subjected to direct amplification ($p < 0.001$) (RFU mean: MT = 6201, MF = 7105, EXT = 13859) (see **Figure 2**). This suggests that DNA extraction is more effective in situations where larger amounts of cellular material are deposited, as the extraction process can recover more DNA from these surfaces. Artifacts such as split and shoulder peaks were observed in the direct amplification profiles, which is consistent with previous studies [7-8,15]. However, despite the presence of these artifacts, true alleles were easily distinguishable from the single-source DNA profiles.

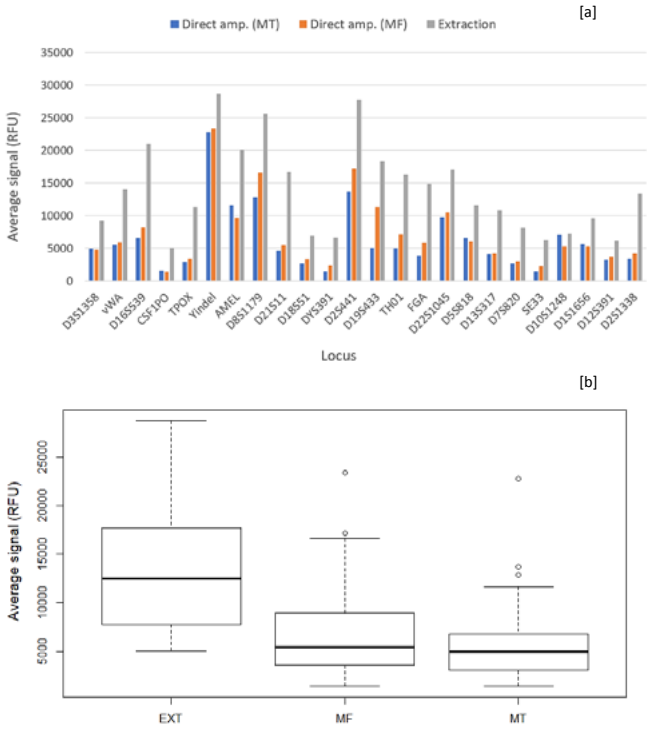


Figure 2 – Comparison of the average signal (RFU) per locus for samples [a] and the mean signal (RFU) per DNA profile for samples [b] (n=144) collected from a large surface area **(5 × 7 cm)**. *T*he figure contrasts samples processed via direct amplification using MicroFLOQ™ Direct swabs (MF) and SceneSafe Fast™ minitape (MT), both amplified with the GlobalFiler™ kit, against samples subjected to extraction (EXT) using the PrepFiler Express BTA™ kit, followed by amplification with the GlobalFiler™ kit.

4. Discussion

The results of this study highlight the differences in DNA profiling outcomes when comparing direct amplification techniques (MT and MF) with traditional DNA extraction, particularly in relation to the size of the deposition area.

For smaller surface areas (2.5 × 3.5 cm), direct amplification using MT and MF proved to be more efficient than extraction methods, yielding higher RFU values. This can be attributed to the high concentration of epithelial cells and cellular material on smaller surfaces, which allows MT and MF to more effectively collect Touch DNA without the need for extraction. The ability of direct amplification methods to bypass the extraction step also offers a more rapid workflow, which is advantageous in forensic applications. However, the appearance of artifacts such as split and shoulder peaks should be noted, although they do not hinder the interpretation of single-source profiles.

In contrast, for larger surface areas (5 × 7 cm), traditional DNA extraction outperformed direct amplification, yielding significantly higher RFU values. This indicates that extraction methods are more effective when larger quantities of DNA are available for recovery. The larger surface area allows for the collection of more cells, and the extraction process is better suited to isolate and concentrate these larger amounts of DNA. The design of the MicroFLOQ™ Direct swabs (MF) may limit the amount of trace DNA that can be collected from larger surfaces, as their smaller tip size might leave some DNA uncollected [7].

5. Conclusion

In summary, this study highlights the effectiveness of various DNA collection methods for Touch DNA profiling across different surface sizes. Direct amplification using MicroFLOQ™ Direct swabs (MF) and SceneSafe Fast™ minitape (K545) (MT) proved superior for smaller surface areas, while DNA extraction demonstrated distinct advantages for larger surfaces. A hybrid approach, utilizing MF or MT for direct amplification in combination with cotton or nylon swabs for extraction, maximizes DNA recovery. These findings emphasize the importance of selecting the appropriate method based on surface size, thereby improving the efficiency and accuracy of forensic DNA analysis.

6. Acknowledgments

This study was approved by the General Department of Forensic Science and Criminology, Dubai Police. Gratitude is extended to colleagues in the Biology and DNA Section for their support.

7. Conflict of interest statement

The authors declare no conflicts of interest.

8. References

[1] S.K. Alketbi, The affecting factors of touch DNA, J. Forensic Res. 9 (3) (2018) 424.

[2] S.K. Alketbi, Analysis of Touch DNA. Doctoral thesis, University of Central Lancashire, (2023). Available at: https://clok.uclan.ac.uk/46154/

[3] S.K. Alketbi, The role of DNA in forensic science: A comprehensive review, Int. J. Sci. Res. Arch. 09 (02) (2023) 814-829.

[4] S.K. Alketbi, W. Goodwin, The effect of surface type, collection, and extraction methods on touch DNA, Forensic Sci. Int. Genet. Suppl. Ser. 7(1) (2019) 704-706.

[5] S.K. Alketbi, W. Goodwin, The effect of time and environmental conditions on Touch DNA, Forensic Sci. Int. Genet. Suppl. Ser. 7(1) (2019) 701-703.

[6] J.L. Barta, et al., One of the key characteristics of ancient DNA, low copy number, may be a product of its extraction, J. Archaeol. Sci. 46 (2014) 281-289.

[7] S.K. Alketbi, An innovative solution to collect Touch DNA for direct amplification, J. Forensic Sci. Crim. Investig. 16(1) (2022) 555928.

[8] S.K. Alketbi, W. Goodwin, Collection methods for Touch DNA direct amplification, J. Forensic Legal Investig. Sci. 9 (2023) 72.

[9] S.K. Alketbi, W. Goodwin, Evaluation of microFLOQ™ Direct Swab for Touch DNA recovery, Forensic Legal Investig. Sci. 10 (2024) 093.

[10] B. Stoop, et al., Touch DNA sampling with SceneSafe Fast™ minitapes, Legal Med. 29 (2017) 68-71.

[11] S.K. Alketbi, W. Goodwin, The impact of deposition area and time on Touch DNA collected from fabric, Forensic Sci. Int. Genet. Suppl. Ser. 8(1) (2022) 45-47.

[12] S.K. Alketbi, The impact of collection method on Touch DNA collected from fabric, J. Forensic Sci. Crim. Investig. 15(5) (2022) 555922.

[13] S.K. Alketbi, W. Goodwin, The impact of area size and fabric type on Touch DNA collected from fabric, J. Forensic Sci. Crim. Investig. 16(1) (2022) 555926.

[14] S.K. Alketbi, S. Alsoofi, Dual recovery of DNA and fingerprints using minitapes, J. Forensic Sci. Crim. Investig. 16(1) (2022) 555929.

[15] A. Ambers, et al., Direct PCR amplification of DNA from human bloodstains, saliva, and touch samples collected with microFLOQ® swabs, Forensic Sci. Int. Genet. 32 (2018) 80-87.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1043-1049
DOI: https://dx.doi.org/10.15304/cc.2025.1869

An alternative method: a simple postmortem DNA sampling and typing for efficient DVI and missing persons identification

Dragana Zgonjanin[1,2*], Rashed Alghafri[3], Reem Lootah[3], Reem Almheiri[3], Péter L. Nagy[4,5], Stojan Petković[1,2]

[1]*Institute of Forensic Medicine, Clinical Center of Vojvodina, Novi Sad, Serbia*

[2]*Faculty of Medicine, University of Novi Sad, Novi Sad, Serbia*

[3]*General Department of Forensic Sciences and Criminology, Dubai Police G.H.Q., Dubai, United Arab Emirates*

[4]*Department of Pathology, Laboratory of Personalized Genomic Medicine, Columbia University, New York, NY, USA*

[5]*Praxis Genomics LLC, Atlanta, GA, USA*

** Corresponding author at: Institute of Forensic Medicine, Clinical Center of Vojvodina, Novi Sad, Serbia, Tel.: +381 63 590 686,*

E-mail address: dragana.zgonjanin.bosic@gmail.com *(D. Zgonjanin)*

dragana.zgonjanin-bosic@mf.uns.ac.rs *(D. Zgonjanin)*

Abstract

The identification of human remains belonging to missing persons is one of the main challenges for the forensic genetics DNA analysis from unidentified human remains. It consists of many steps, but the extraction and optimized recovery of DNA is of paramount importance in this process. Since the beginning in 2003, our laboratory has focused exclusively on STR DNA from a bone. FTA card is unfortunately rarely used to collect samples from decomposed bodies. The number of samples submitted for body identifications to our laboratory by the law enforcement has increased in the last few years. Hence the ability to efficiently and rapidly extract DNA from a variety of post-mortem sample types and perform matching with family or direct reference samples in a short period of time is one of the main challenges that our laboratory has been facing. The purpose of this paper is to provide sampling and DNA typing guidelines that would make it both easier and more cost effective than methods currently used in such cases (e.g. DNA typing from bones/teeth). Rib cartilage samples were used for a DNA identification of 10 bodies that were autopsied at the Institute of Forensic Medicine. Extraction of DNA was performed using QIAamp DNA Micro Kit (QIAGEN), 5% Chelex 100 (Bio-Rad) and InstaGene™ Matrix

(Bio-Rad). DNA profiles were generated from DNA extracts using GlobalFiler™ PCR Amplification kit (Thermo Fisher Scientific). It was of particular importance to obtain complete DNA profiles irrespective of environmental conditions from which the bodies were recovered. Our work has shown that with post-mortem DNA sampling and simple extraction protocol, using more sensitive and robust new amplification kits we were able to overcome the challenges associated with processing compromised skeletal remains and to perform faster, easier and more cost effective analysis of these types of samples.

Keywords
Skeletal remains, DNA typing, Human identification, Rib cartilage samples.

1. Introduction

Various forensic techniques are used to identify a human corpse, depending on the circumstances and the state of remains. When other human identification forensic techniques (e.g. dactyloscopy, anthropology, odontology and medico-legal examinations) provide limited information, or to support further or refute potential associations, DNA typing can be extremely valuable. DNA, in theory, can be recovered essentially from any tissue (e.g. soft, degraded tissues, bones, teeth and hairs) [1–4].

In our long-standing work with human identification cases using bone and tooth samples, following two different protocols: phenol chloroform isoamyl alcohol (PCIA) organic extraction method [5] and PrepFiler® BTA Forensic DNA Extraction Kit (Applied Biosystems) [6-9], modifications to both extraction methods have been introduced in order to improve our laboratory success rate with identification of these skeletal remains [5,6]. Zgonjanin et al. [5] have reported modifications to their DNA extraction protocol that have allowed them to successfully generate DNA profiles from burned bodies.

Our aim was to avoid extreme delays and laborious processing of the bone and tooth samples (including mechanical and chemical cleaning, cutting and powdering followed by labor-intensive demineralization, and finally DNA extraction with hazardous chemicals) in order to obtain DNA profiles in human identification cases as quickly as possible. To combat degradation and environmental factors associated with different samples, the scientist can optimize sample selection. Once sample selections have been optimized, the laboratory can increase success rates through the enhancement of extraction methods to guarantee complete cell lysis and amplification methods to combat degradation and inhibition.

An in-house developed method with appropriate sample selection from rib cartilage from decomposed bodies, using simplified manual protocols including only one pipetting step with Chelex-based methods and the simple QIAamp DNA Micro procedure, which are both highly suited for simultaneous processing of multiple samples, yields pure DNA in less than 90 minutes. Compared to our former method [5-9] with bones/teeth, this reduced the amount of manual labor as well as the time needed for extraction from two days to just 90 minutes without compromising the quality of the obtained DNA profiles.

2. Material studied, methods, techniques

2.1. Human DNA Samples

This study analysed 10 rib cartilage samples from decomposed bodies, submitted for routine casework body identifications by law enforcement. In a sterile environment, a tissue slicewas excised from each preserved rib cartilage sample using DNA-free scalpel/scissors. The clean and cut bone fragments were washed and air dried. The rib cartilage was cut into pieces of 2x2 mm, and for each individual analysis, 2 to 3 pieces (about 30 mg) were used. Any mass of tissue can be excised but it may be wise to leave some for a repeat analysis. For example, we removed approximately 30 mg (roughly 20% of our total collected tissue).

2.2. DNA Extraction

Total DNA was isolated from rib cartilage samples following three different protocols: Chelex method (5% Chelex 100, Bio-Rad), InstaGene™ Matrix (Bio-Rad) and QIAamp DNA Micro Kit (QIAGEN). The two developed DNA extraction protocols based on common Chelex method consist of the following steps: 170 µl 5% Chelex solution and 20 µl of 20 mg/ml proteinase K; also 200 µl of Instagene Matrix 20 µl of 20 mg/ml proteinase K. The samples were vortexing on high for 10 seconds, followed by two incubation steps in a heating incubator (56°C for 45 min and 100°C for 8 min) with vortexing in between. In the third applied protocol DNA was purified from decomposed rib cartilage samples using the QIAamp® DNA Micro Kit [10].

2.3. PCR amplification and typing

DNA was quantified with an ABI Prism® 7500 Sequence Detection System (Applied Biosystems) using Quantifiler™ Human DNA Quantification kit.

Amplifications were performed on the GeneAmp PCR System 9700 Gold Plate (Applied Biosystems) using the AmpFLSTR® Identifiler® Plus kit (Applied Biosystems), GlobalFiler™ PCR Amplification kit (Thermo Fisher Scientific), and AmpFlSTR® Yfiler® Plus PCR Amplification Kits (Applied Biosystems) following the manufacturers' protocols. Amplified products are separated and detected on ABI 3500 Genetic Analyzer (Applied Biosystems).

3. Results

Ten DNA extractions were completed on unidentified human remains/decomposed bodies from routine casework using rib cartilage samples and the results are presented in Table 1. Three different groups of bones were selected based on exposure of skeletal remains to different environmental influences which may have led to degradation of DNA: burned bodies (3 samples), remains recovered from water (3 samples) and remains recovered from fields (4 samples) (Table 1). All samples, from each amplification kits produced full STR profiles beside on extraction method are used (Table 1).

We have also successfully extracted DNA using the following commercial kit QIAamp® DNA Micro Kit (QIAGEN) as and Chelex based methods Chelex method (5% Chelex 100, Bio-Rad), InstaGene™ Matrix (Bio-Rad). All three extraction protocols provided (comparable amounts of) DNA yields and complete STR profiles of equivalent quality (Table 1).

Family reference samples accompanying the skeletal remains are usually obtained from first-degree relatives or multiple direct references/other belongings likely to have accumulated DNA (e.g., tooth brushes, electric and manual razors, hair brushes and combs, pillowcases, eye glasses, clothes, jeans etc) (Table 1). When comparing genetic profiles, we matched 8 out of the 10 skeletal remains analyzed to accompanying reference sample with high confidence of correct identification (probability from 99.9% to 99.999999%). In the last two cases, direct reference samples were used that belonged to victims/missing persons (in case #4, pillowcases and jeans; in case #7, toothbrush and manual razors) (Table 1).

Table 1

Nuclear DNA quantity; the efficiency of STR and Y-STR typing expressed as the number of successfully typed STRs in 10 rib cartilage samples from decomposed bodies, submitted for routine human identification cases.

4. Discussion

Through in-house protocols we have shown that the method is fit-for-purpose for application in casework, as it provides high DNA yields and amplifiability, as well as good reproducibility, easier and more cost effective. After implementation in casework, the proportion of extracts with DNA concentrations above 0.10 ng/mL increased to 45% using rib cartilage samples, compared to 45% extraction from bone samples used to collect samples from decomposed bodies, only 10% of human identification cases used FTA cards. Apart from providing equal DNA yields compared with the previous method working with bone samples, the introduction of the developed simple protocol using rib cartilage samples also reduced the amount of manual labour and increase the potential throughput for casework at the laboratory. Generally, simplified manual protocols can serve as a cost-effective alternative to sophisticated automation solutions when the aim is to enable high-throughput DNA extraction of unidentified human remains/decomposed bodies in casework.

5. Conclusion

This study was aimed at improving identification techniques based on the analysis of genomic DNA. The results of this research suggest that depending on the state of preservation, different tissue types should be collected, allowing for bone samples to be replaced with rib cartilage. The STR profiles generated when extracting DNA using the Chelex-based methods were comparable to those generated using the commercial QIAamp® DNA Micro Kit procedure. When dealing with decomposed bodies/skeletal remains, all three kits can be used very successfully without any changes to the manufacturers' PCR amplification protocols.

Our results have shown that simple extraction techniques, together with appropriate sample selections such as rib cartilage from decomposed bodies and simplified manual protocols, have allowed us to overcome the challenges associated with processing compromised skeletal remains and have led to the identification of more missing individuals.

The method was implemeted at the Institute of Forensic Medicine, Clinical Center of Vojvodina in January 2020. During the first year more than half of the human identification requests were extracted using the method from rib cartilage samples, generating DNA yields and STR profile quality equivalent to previously used methods.

6. Conflict of interest statement

None.

7. References

[1] E. Hagelberg, B. Sykes, R. Hedges, Ancient bone DNA amplified, Nature 342 (1989) 485.

[2] E. Hagelberg, I.C. Gray, A.J. Jeffreys, Identification of the skeletal remains of a murder victim by DNA analysis, Nature 352 (1991) 427–429.

[3] P. Gill, P.L. Ivanov, C. Kimpton, et al., Identification of the remains of the Romanov family by DNA analysis, Nat. Genet. 6 (1994) 130–135.

[4] M.M. Holland, D.L. Fisher, L.G. Mitchell, et al., Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War, J. Forensic Sci. 38 (1993) 542–553.

[5] D. Zgonjanin, S. Petković, M. Maletin, et al., Case report: DNA identification of burned skeletal remains, Forensic Sci. Int.: Genet. Suppl. Ser. 5 (2015) e444–e446.

[6] D. Zgonjanin, M. Antov, R. Alghafri, et al., DNA analysis from human skeletal remains in forensic casework, Forensic Sci. Int.: Genet. Suppl. Ser. 6 (2017) e342–e345.

[7] D. Zgonjanin, D. Nedić, R. Alghafri, et al., DNA typing from skeletal remains using GlobalFiler™ PCR amplification and Investigator® 24plex QS kits, Forensic Sci, Int. Genet. Suppl. Ser. 7 (2019), e50-e52.

[8] D. Zgonjanin, E. Almohammed, D. Nedić, et al., Importance of DNA analysis for identification and confirmation of human remains, following a forensic autopsy, Forensic Sci, Int. Genet. Suppl. Ser. 2019. 7 (2019), e53-e55.

[9] E. Almohammed, D. Zgonjanin, A. Iyengar, et al., A study of degraded skeletal samples using ForenSeq DNA Signature™ Kit, Forensic Sci. Int.: Genet. Suppl. Ser. 6 (2017) e410–e412.

[10] QIAamp® DNA Micro Handbook, QIAGEN, Third Edition December 2014, pp. 1–40.

Table 1. Nuclear DNA quantity; the efficiency of STR and Y-STR typing expressed as the number of successfully typed STRs in 10 rib cartilage samples from decomposed bodies, submitted for routine human identification cases.

Bone sample (Q)/ casework	Environmental exposure	Extraction method	Quantity (ng/µl)	Efficiency of STR typing			Relationship of K to Skeletal Remains/ direct references
				Identifiler® Plus	GlobalFiler™	Yfiler® Plus	
Rib cartilage[a]/Case #1	Buried	Chelex	3.892	16/16	24/24	27/27	son
Rib cartilage[b]/Case #2	Burned	QIAamp® DNA Micro Kit	1.151	16/16	24/24	-	mother
Rib cartilage[a]/Case #3	Water	Chelex	5.256	16/16	24/24	27/27	father
Rib cartilage[a]/Case #4	Buried	QIAamp® DNA Micro Kit	2.112	16/16	24/24	27/27	pillowcase; jeans
Rib cartilage[b]/Case #5	Burned	InstaGene™ Matrix	2.989	16/16	24/24	-	daughter
Rib cartilage[a] / Case# 6	Buried	InstaGene™ Matrix	3.805	16/16	24/24	27/27	brother
Rib cartilage[a] /Case #7	Water	QIAamp® DNA Micro Kit	6.949	16/16	24/24	27/27	toothbrush; razor
Rib cartilage[b]/Case #8	Water	InstaGene™ Matrix	0.778	16/16	24/24	-	mother
Rib cartilage[a] /Case #9	Burned	Chelex	0.794	16/16	24/24	27/27	brother
Rib cartilage[a] /Case #10	Buried	InstaGene™ Matrix	2.928	16/16	24/24	27/27	son

[a] man

[b] woman

Q-questioned; K-known; – not performed

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1051-1059
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Internal Validation of the Investigator 26Plex QS Amplification Kit: a high-throughput multiplex assay for reference and low copy number DNA samples

Paula Cardoso[1*], Armando Serra[1], Vanessa Bogas[1], Filipa Balsa[1], Virginia Lopes[1], Rita Dario[1], Maria João Porto[1], António Amorim[1,2,3], Francisco Corte Real[1,4], Pedro Brito[1]

[1]*Instituto Nacional de Medicina Legal e Ciências Forenses, I.P., Coimbra, Portugal.*

[2]*Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal.*

[3]*REQUIMTE, Analytical Development Group, Lisboa, Portugal.*

[4]*Faculdade de Medicina da Universidade de Coimbra, Coimbra, Portugal.*

** Corresponding author at: paula.l.cardoso@inmlcf.mj.pt*

Abstract

Human singularity provides a foundation for forensic genetics, where the study of Short Tandem Repeats (STRs) is key. Furthermore, the development of polymerase chain reaction (PCR) has revolutionized the field by enabling the amplification of specific DNA markers. The Investigator® 26plex QS Amplification Kit is a highly robust multiplex PCR system, for human identification. In this study, we report the results obtained from the internal validation carried out at the Portuguese National Institute of Legal Medicine and Forensic Sciences. The results showed that the Investigator® 26plex QS kit proved to be a robust tool for the analysis of forensic samples, offering high sensitivity and specificity.

Keywords

Validation, Amplification Kit, Thresholds, STRs.

1. Introduction

The Investigator® 26plex QS Amplification Kit, from QIAGEN, provides reliable and rapid DNA profiles while enabling the multiplex amplification of 24 STRs, 2 Quality Sensors and a gender-determining marker, Amelogenin. The Quality Sensors incorporated in this kit provide insight into the sample quality and the PCR's success while alerting to the presence of inhibitors [1, 2].

Through an extensive internal validation study, this research sought to implement the Investigator® 26plex QS in the laboratory routine of the Forensic Genetic and Biology Service, Central branch (SGBF-C) of the Portuguese National Institute of Legal Medicine and Forensic Sciences. Here we detail the procedures and parameters applied which followed the Scientific Working Group on DNA Analysis Methods (*SWGDAM*) guidelines, as well as report the results obtained throughout [3, 4].

Firstly, it was crucial to establish unique analytical criteria that would be the baseline for the interpretation of the results obtained. To accomplish such, analytical, stochastic, heterozygous imbalance and stutter thresholds were defined. Thereupon, this study intended to gauge the kit's performance, by evaluating its concordance with normalized methodologies while assessing its sensitivity, specificity, robustness, and precision, besides its efficiency in the presence of degraded and/or inhibited samples. Lastly, in favor of optimizing the laboratory workflow, an assay was carried out to test half volume reactions and direct amplification on reference samples.

2. Material studied, methods, techniques

2.1. Overall Procedure

In this study a wide range of sample types were analyzed, which made it possible to acquire robust data pertaining to the performance of the assay. The laboratory methodology applied comprehended: DNA extraction using Prep-n-Go™ Buffer for the study of sensitivity, inhibitor tolerance, mixture samples and procedure optimization; and PrepFiler Express™ Forensic DNA Extraction Kit for the study of specificity and casework samples. Which was followed by quantification with the Quantifiler™ Trio Quantification Kit, and amplification with the Investigator® 26plex QS. Finally, capillary electrophoresis was performed using the Applied Biosystems™ 3500 Genetic Analyzer, and electropherograms were analyzed with using GeneMapper™ ID-X Software v1.6.

2.2. Thresholds

In order to evaluate and standardize the analysis of electropherograms by different professionals, it is crucial to establish unique criteria to guide its interpretation. To determine the analytical threshold (AT) of this methodology, 8 PCR negative controls were used and applied in the 3500 Genetic Analyzer. After the

first injection, six re-injections were performed, totaling 56 negative samples. For every electropherogram obtained, the highest detected RFU intensity in each color panel was recorded. The AT was calculated using Kaiser's approach, $AT = \bar{x} + 3 * \sigma$, where $\bar{x}$ refers to the average of the highest RFU intensity for each sample and σ to the associated standard deviation. This procedure was conducted for an injection time of 10 and 40 seconds [4, 5].

To establish the stochastic threshold the SWGDAM (2010) recommended approach was used, where the peak height ratio across multiple loci is assessed in a dilution series. To carry this out, two known male samples with a high number of heterozygosity were selected. These samples were then quantified and diluted to an approximate concentration of 2.5 ng/µL. Thereupon 6 serial 1:2 dilutions for each sample using nuclease-free water were prepared and analyzed at 10 and 40 seconds of injection time [4, 6].

In the early 2000s, forensic labs used a standardized 15% stutter threshold (ST) across all systems. After SWGDAM's 2010 recommendations, thresholds became system-specific due to variability among STRs. For this essay, 150 electropherograms were analyzed, stutter proportions were calculated, and ST values were determined using the formula $ST = \bar{x} + 3\sigma$. Finally, the values obtained were compared with those provided by the kit manufacturer to ensure accuracy under SGBF-C conditions [4].

To determine the intra-locus peak height ratios (PHR), 125 electropherograms were analyzed and the GHEPMIX Commission's method was employed. The intra-locus peak height ratios, calculated as the RFU height of the smaller allele divided by the larger allele, was used to define PHR. The threshold was then determined using the formula $LD = \bar{x} + 3\sigma$, where $\bar{x}$ is the average ratio and σ is the standard deviation [7].

2.3. Specificity

The specificity of this methodology was evaluated by analysing extracted samples from different species of vertebrate animals (*Equus caballus*, *Canis lupus familiaris*, *Sus scrofa domesticus*, *Cervus elaphus*, *Equus asinus*, *Ovis aries*, *Mustela putorius furo* and *Felis catus*), followed by amplification with the Investigator® 26Plex QS with an injection time in the 3500 Genetic Analyzer of 10 seconds.

2.4. Sensitivity

According to Qiagen's manual, the recommended DNA input for optimal results is 0.5 ng, in this study a sample with a known concentration of 2,5 ng/µL was

diluted to the expected concentrations of 0.5 ng/µL, 0.1 ng/µL, 0.05 ng/µL, 0.029 ng/µL, 0.02 ng/µL, 0.013 ng/µL, and 0.01 ng/µL. This samples were then amplified in order to observe the minimum DNA quantities that still produce a valid genetic profile. This study was performed for 10 and 40 seconds injection time on the 3500 Genetic Analyser.

2.5. Inhibition Tolerance

To assess the Investigator 26Plex QS' tolerance to inhibitors presence during amplification, samples were prepared using the 2800M Control DNA (Promega) and five different inhibitors – (1) indigo carmine, (2) ethanol, (3) foliage, (4) clay soil, and (5) hemic soil - at varying Inhibitor vs DNA ratios, namely 1:10, 1:1, and 10:1.

To prepare the inhibitors solutions: (1) a 12 mM indigo carmine solution was prepared using indigo carmine (Sigma-Aldrich) and nuclease-free water, (2) 70% ethyl alcohol was used directly and (3, 4, 5) approximately 1 g of each substrate - foliage, clay and hemic soil - was macerated in 2 mL of nuclease-free water until homogeneous, the mixture was then centrifuged, and the supernatant removed.

After agitating the solutions for 3 hours, samples were amplified with DNA inputs of 1 ng, 0.5 ng, and 0.1 ng for the respective Inhibitor vs DNA ratios - 1:10, 1:1, and 10:1. In total 18 samples, 15 with inhibitors and 3 control samples with nuclease-free water, were amplified with the Investigator® 26Plex QS with an injection time in the 3500 Genetic Analyzer of 10 seconds.

2.6. Mixture samples

Five extract samples (A, B, C, D, E) quantified and diluted to a final concentration of 0.5 ng/µL were selected and combined to create five different mixtures (A+B, C+D, B+D, B+E, and D+A) at concentrations of 1:50, 1:20, 1:10, 1:1, 10:1, 20:1, and 50:1, resulting in a total of 35 controlled mixtures analysed an injection time of 10 seconds.

2.7. Procedure optimization

The recommended final reaction volume by the manufacturer is 25 µL and numerous protocols are available for different sample types. Notwithstanding, in order to optimize the reaction volume and uniformize the laboratory procedure independently of sample aetiology, a protocol for half-volume reaction for both extracted samples and direct amplification was tested.

For that 24 extracted samples were amplified following both the manufacturer's protocol and a half-volume reaction protocol (12.5 µL final volume). This latter method was carried out in duplicate for 25 and 26 cycles, in order to better understanding of which would be more suitable for the laboratory routine.

Direct amplification of 12 blood stains with a punch size of 0.5 mm were analysed using the same methodology.

2.8. Forensic casework

In order to evaluate the Investigator® 26Plex QS' performance and ability to obtain consistent results, 5 samples from proficiency collaborative exercises organized by the Spanish and Portuguese-Speaking Working Group of the International Society for Forensic Genetics and 5 forensic casework samples were selected and amplification was performed according to manufacturer's protocol.

3. Results

3.1. Thresholds

Thresholds are crucial for standardization in electropherogram analysis. The Investigator® 26Plex QS kit exhibited thresholds that are on par with those described in the literature.

The analytical threshold, value above which a signal can be considered a true allele, was defined as 50 RFU and 100 RFU, for an injection time of 10 seconds and 40 seconds, respectively. The stochastic threshold, value above which allelic dropout is excluded, was determined as 200 RFU and 400 RFU for an injection time of 10- and 40 seconds, respectively. Furthermore, the intra-locus peak height ratio was set as 60%, a consensual value in the scientific community.

This study also defined the stutter threshold for each genetic marker, confirming the adequacy of the values provided by the manufacturer. The provided thresholds were implemented for all markers except for STRs D10S1248 and D16S359, which required distinct thresholds of 4–16% and 12%, respectively, to reduce detected artifacts.

3.2. Specificity

The analysis of several different non-human animal specifies did not provide the amplification of a single allele.

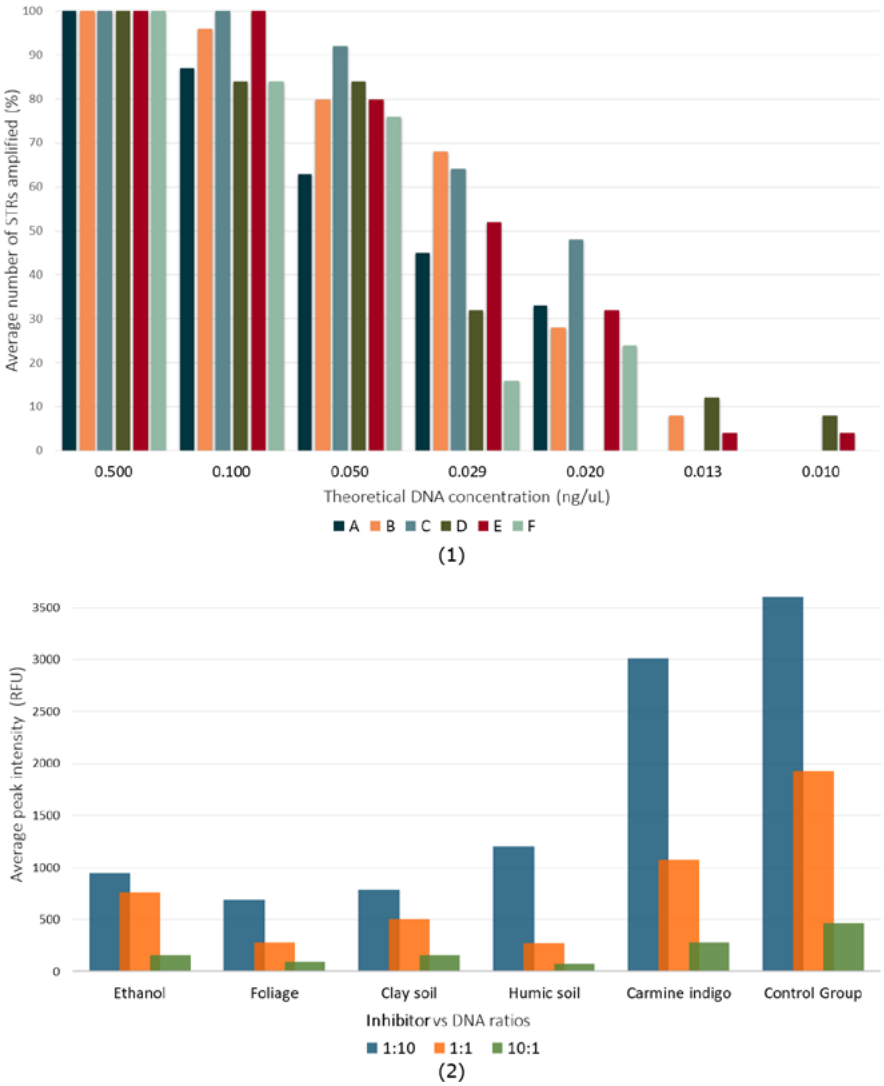


Figure 1. Sensitivity and inhibition tolerance results: (1) Investigator® 26Plex QS kit's sensitivity essay in 6 samples (A to F) and (2) the effects of various inhibitors at different ratios (Inhibitor vs DNA).

3.3. Sensitivity

The results of this study showed that the optimal DNA input of this amplification study is adequate for robust analysis. It also showed that it is possible

to attain complete profiles up to a0.1 ng - low template DNA - at which point the detection of STRs is compromised, with preferential amplification of smaller markers (Figure 1 - 1). The application in the 3500 Genetic Analyzer with 40 seconds of injection time was more informative, allowing for the characterization of more alleles.

3.4. Inhibition tolerance

The effect of DNA inhibition on the performance of the Investigator 26Plex QS Kit was not significant in the Inhibitor vs DNA ratio of 1:10 and 1:1.

Furthermore, this study highlighted that the amplification's efficiency was only affected (decreased allelic intensity and allelic drop-out), in those samples where the inhibitor proportion was higher whilst the DNA amount was minimal – 10:1 Inhibitor vs DNA (Figure 1 - 2). It is important to note that this proportion is hypothetical, since the casework samples frequently studied have considerately less inhibition present during amplification.

3.5. Mixture samples

Complete profiles of both contributors were obtained for ratios of 1:10, 1:1 and 10:1. However, the other mixture proportions analysed, namely 1:50, 1:20, 20:1 and 50:1, showed an allelic drop-out of the minor contributor. Nevertheless, since a mixed DNA profile arises from at least two contributors and by definition in the SGBF-C a profile with 3 or more alleles in two or more loci and a recurrent allelic imbalance is considered a mixture, this amplification kit detected mixtures up to proportions of 50:1 and 1:50.

3.6. Procedure optimization

The changes proposed to the manufacturer's protocol, namely the half-volume reaction methodology for both extract samples and for direct amplification of blood stains has provided valid, concordant and robust results.

3.7. Forensic caseworks

All of the tested casework samples amplified with this assay generated full profiles and all genotypes were concordant with those obtained with the amplification kits currently implemented in the SGBF-C. Nevertheless, the Investigator® 26Plex QS produced fewer artefacts in the most challenging samples with a greater precision.

4. Discussion

Through this work, it was possible to characterize the main advantages of the Investigator® 26plex QS kit, as well as its limitations. The results inferred a high human specificity, robustness, and sensitivity, while producing concordant and reproducible results independently of the operator, equipment or methodology used - full-volume reaction, half-volume reaction and direct amplification.

The validation data demonstrated that this system produces full reliable profiles in the presence of minute DNA quantities – 0.1 ng - and allows for the identification of a minor contributor in a mixture up to a 1:50 ratio, where the minor contributor's input was only 0.02 ng.

The analysis of the sensitivity and mixtures parameters, showed that a 40s injection time in the 3500 Genetic Analyzer provides more informative results for samples with small DNA quantities. Therefore, when quantification indicates that the maximum DNA input in the PCR reaction is below 0.5 ng, this methodology should be immediately applied.

This project also demonstrated that, for optimal DNA concentrations, this amplification kit can overcome the presence of inhibitors in a sample and produce complete genetic profiles. Moreover, since the Inhibitor vs DNA ratio used in this analysis is beyond the level expected and usually observed in forensic casework, even with low copy number DNA, the Investigator® 26Plex QS's performance should not compromised.

5. Conclusion

In summary, through this internal validation studies, numerous samples from different substrates were amplified using distinct methodologies – full-volume reaction, half-volume reaction and direct amplification – and all provided concordant results with high precision, sensitivity and robustness. Therefore, the Investigator® 26Plex amplification kit is suitable for the amplification and characterization of forensic samples.

6. Acknowledgments

The authors would like to gratefully acknowledge Qiagen for providing us one amplification kit.

7. Conflict of interest statement

The authors declare that they have no competing interests.

8. References

[1] QIAGEN. Developmental Validation of the Investigator® 26plex QS Kit, 2020. https://www.qiagen.com/us/resources/download.aspx?id=96358d36-7418-42e4-bd41-0eaa0b0eae01&lang=en

[2] M. Vraneš, M. König, S. Cornelius, A. Kohns, K. Elliott, A. Prochnow, M. Scherer. Development of the Investigator® 26plex QS Kit: A New multiplex PCR Kit for Global STR analysis. Forensic Sci. Int. Genet, 7 (2019), 778-779.

[3] SWGDAM. Scientific Working Group on DNA Analysis Methods: Validation Guidelines for DNA Analysis Methods, 2016. https://www.swgdam.org/_files/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf

[4] ENFSI. Guidelines for the single laboratory: Validation of Instrumental and Human Based Methods in Forensic Science, Reference Code: QCC-VAL-002, 2014. https://enfsi.eu/wp-content/uploads/2017/06/Guidance-QCC-VAL-002.pdf

[5] J. M. Butler. Advanced Topics in Forensic DNA Typing: Methodology. Academic Press, Elsevier, 2011.

[6] S. Bell and J. M. Butler. Understanding Forensic DNA. Cambridge University Press, 2022. https://doi.org/10.1017/9781009043311

[7] M. Crespillo, J. A. Luque, M. R. Paredes and P. A. Barrio. Criterios mínimos recomendados para la Aceptación y Evaluación de Perfiles Mezclas. Recomendaciones Comisión GHEPMIX, 2012. https://doi.org/10.13140/RG.2.2.14617.72800

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1061-1066
DOI: https://dx.doi.org/10.15304/cc.2025.1869

High Performance PCR Assay System Capable of Cleaning Carry-Over Contamination

Dongke Fu[1]

[1]*NUHIGH Biotechnologies Co., Ltd. Suzhou, China*

Abstract

The enhanced sensitivity of STR multiplex PCR kits has led to increased sample throughput in forensic laboratories, accompanied by a rise in DNA contamination cases. Contamination undermines test accuracy, misdirects investigations, and delays case resolutions. Uracil-N-glycosylase (UNG), a DNA repair enzyme, is pivotal in addressing carry-over contamination (PCOC) in PCR. The PCOC system, combining UNG and dUTP, effectively mitigates aerosol contamination of PCR products. We developed a 25-loci STR multiplex PCR kit incorporating this system, featuring a 50°C pre-amplification incubation step that eliminates carry-over contamination. Our kit maintains high sensitivity, equivalent to conventional systems, and exhibits significant tolerance to inhibitors such as EDTA, indigo, hematin, humic acid, and collagen. It also shows minimal cross-reactivity with non-human DNA in environmental samples.

Keywords

dUTP/UNG, carry-over contamination, STR, forensic.

1. Introduction

DNA detection is integral to criminal investigations, enabling reliable forensic evidence generation. However, increased sensitivity in DNA testing has heightened contamination risks. A study by the Swiss police forensic department reported 552 contamination cases between 2011 and 2015 [1]. Contamination can mislead investigations and damage the credibility of the judicial system [2]. Thus, minimizing contamination risk is essential.

Forensic DNA laboratories enforce strict protocols to prevent carry-over contamination from PCR products and allelic ladder aerosols. UNG, a DNA repair enzyme, plays a critical role in preventing contamination. In the 1990s, Longo

introduced the dUTP-UNG system, which effectively reduces PCR product contamination and aerosol transmission [3]. However, traditional DNA polymerase is inefficient at incorporating dUTP in high-multiplex STR reactions, particularly for loci with high A/T content, such as D22S1045. We developed a high-performance PCOC system with a 25-loci STR kit, using an engineered Taq DNA polymerase that incorporates dUTP more efficiently. The system utilizes a heat-inactivated UNG enzyme [4, 5], making it more reliable for forensic DNA applications.

The dUTP/UNG system replaces dTTP with dUTP in PCR. UNG hydrolyzes dUTP-containing strands at 50°C, preventing contamination without affecting the PCR product from genomic DNA. At 95°C, UNG is inactivated, ensuring no interference with amplification (Figure 1).

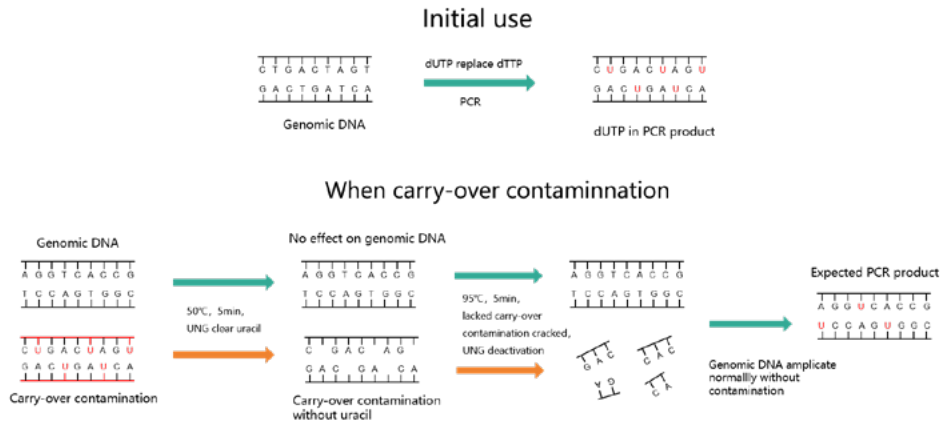


Figure 1. The working principle of the PCOC PCR system

This study highlights a dUTP/UNG forensic STR kit with exceptional contamination control, high sensitivity, and strong inhibitor tolerance.

2. Material studied, methods, techniques

The PCOC STR multiplex PCR kit used in this study includes 23 autosomal STR loci, 1 sex-identification locus, and 1 Y-indel (Figure 2). The dUTP/UNG system PCR Mix, provided by NuHigh Biotechnologies Co. Ltd., contains dATP, dGTP, dCTP, dUTP, a mutant Taq polymerase, UNG, and other reagents.

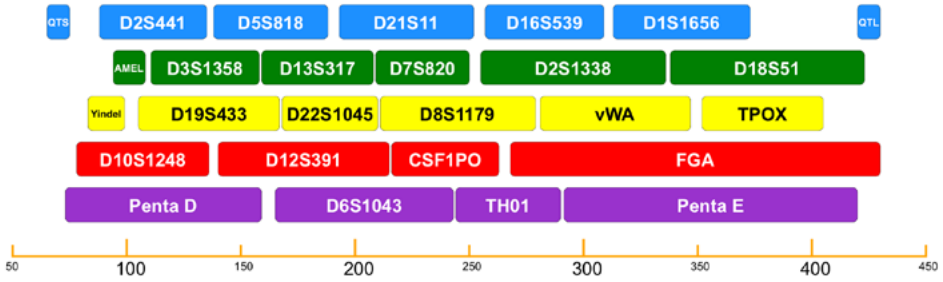


Figure 2. Marker positioning of the PCOC system

3. Results

To evaluate contamination control, simulated contamination (1 ng 9948 DNA diluted 1:10,000,000) was tested. In standard PCR, all alleles were detected with an average peak height of 600 copies (2 ng human DNA) (Figure 3A). In contrast, the PCOC system produced negative readings, confirming complete contamination removal (Figure 3B).

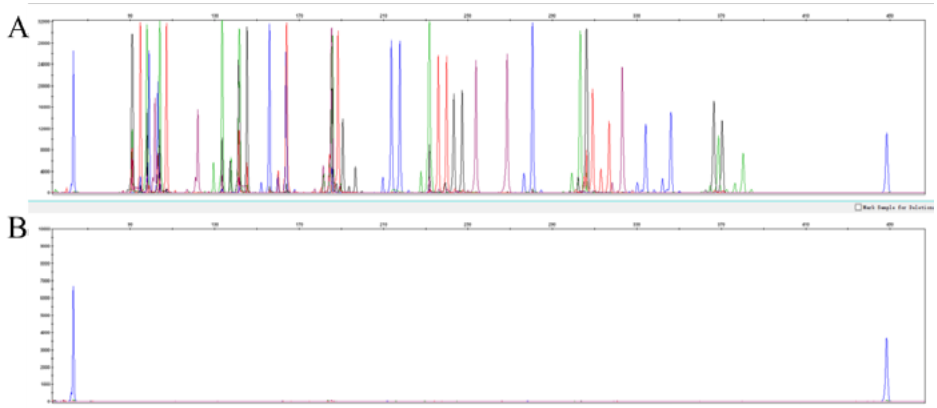


Figure 3. Comparison of PCOC system and standard PCR system

Mixed alleles were detected in the normal PCR system using contaminants and 9947A genomic DNA (Figure 4A), but no additional alleles were observed in the PCOC system (Figure 4B).

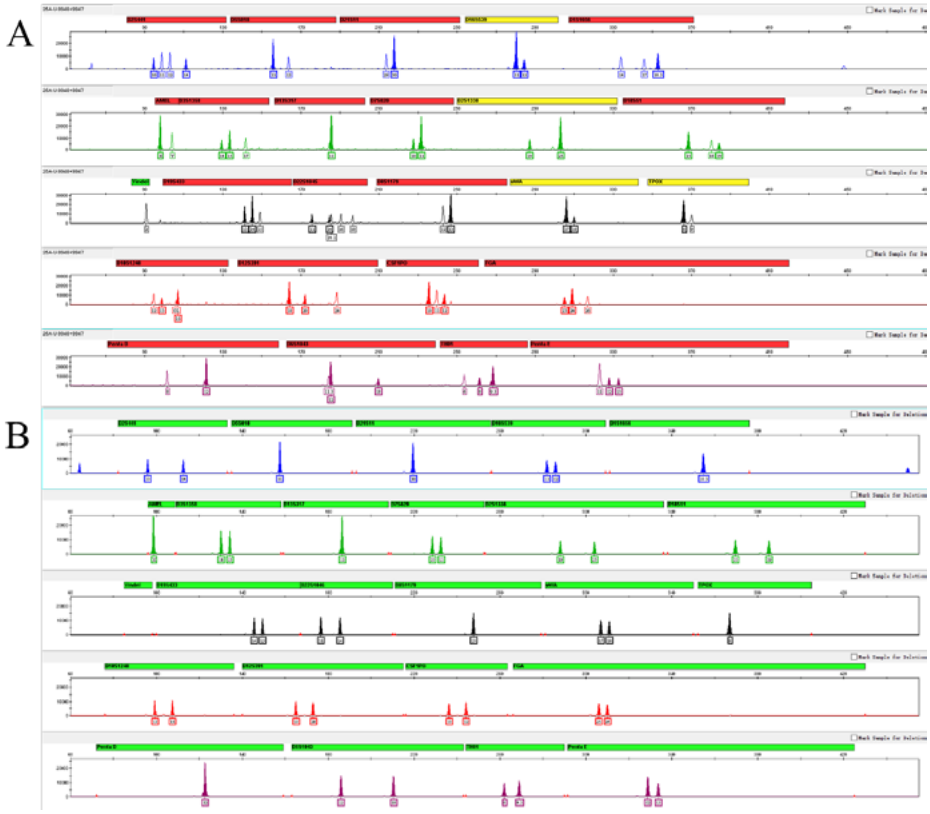


Figure 4. Electropherogram of contaminant and 9947A genomic DNA in standard (A) and PCOC (B) PCR systems.

In sensitivity tests, the PCOC system generated complete STR profiles from as little as 50 pg of human DNA and partial profiles at lower DNA levels (Figure 5A). The sensitivity of the PCOC system was comparable to the conventional PCR system using dTTP. It also demonstrated strong resistance to common forensic inhibitors, including collagen, EDTA, indigo, hematin, and humic acid (Figure 5B).

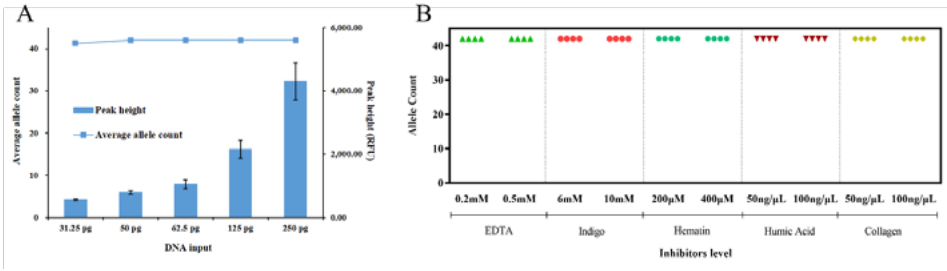


Figure 5. (A) Sensitivity and (B) inhibitor tolerance of the PCOC PCR system.

Species specificity testing showed minimal cross-reactivity with non-human DNA, as no alleles were detected in animal or microorganism samples (including pig, horse, cat, chicken, cow, dog, duck, fish, rabbit, rat, and sheep, as well as Escherichia coli, Candida albicans, Lactobacillus delbrueckii, and others).

4. Discussion

The sensitivity of forensic DNA testing has increased, but even trace amounts of contamination can lead to false positives. UNG-based systems are well-established in in vitro diagnostics but have limited use in forensics. Our PCOC kit offers superior contamination control, with a detection limit of 50 pg, and exhibits high resistance to forensic inhibitors.

5. Conclusion

The PCOC multiplex STR PCR system we developed is highly effective at cleaning carry-over contamination, ensuring accurate DNA detection even in the presence of contamination. The system matches the sensitivity of traditional PCR and is robust against common forensic inhibitors.

6. Acknowledgments

We thank all people involved in sample collection and lab work.

7. Conflict of interest statement

None.

8. References

[1] P. Basset, V. Castella, Lessons from a study of DNA contaminations from police services and forensic laboratories in Switzerland, J. Forensic Sci. Int. Genet. 33 (2017) 147-154.

[2] A. Kloosterman, M. Sjerps, A. Quak, Error rates in forensic DNA analysis: definition, numbers, impact and communication, J. Forensic Sci. Int. Genet. 12 (2014) 77-85.

[3] M. C. Longo, M. S. Berninger, J. L. Hartley, Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. Gene. 93 (1) (1990) 125-128.

[4] J. L. Hartley and A. Rashtchian, Dealing with contamination: enzymatic control of carryover contamination in PCR. PCR Methods Appl. 3 (2) (1993) S10-14.

[5] K. E. Pierce and L. J. Wangh, Effectiveness and limitations of uracil-DNA glycosylases in sensitive real-time PCR assays. Biotechniques. 36 (1) (2004) 44-46.

[6] D. Andersson, D. Svec, C. Pedersen, J. R. Henriksen, Preamplification with dUTP and Cod UNG Enables Elimination of Contaminating Amplicons. Int. J Mol. Sci. 19 (10) (2018) 3185.

[7] E. W. Taggart, K. C. Carroll, C. L. Byington, Use of heat labile UNG in an RT-PCR assay for enterovirus detection. J Virol. Methods. 105 (1) (2002) 57-65.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1067-1072
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Comparative study for the effect of different vectors and pretreatment methods on DNA typing efficiency

Qiannan Xu[1,2,3], Zhe Zheng[1,2,3], Wei Feng[1,2,3], Xiandun Zhai[1,2,3*]

[1]*School of Basic Medicine and Forensic Medicine, Henan University of Science and Technology, Luoyang, 471023, PR China.*

[2]*Institute of Medical Aspects of Specific Environments, Henan University of Science and Technology, Luoyang, 471023, PR China.*

[3]*Forensic Medicine Identification Center, Henan University of Science and Technology, Luoyang, 471023, PR China.*

** Corresponding author at: School of Basic Medicine and Forensic Medicine, Henan University of Science and Technology, Luoyang, 471023, PR China.*

Email: xiandunzhai@163.com; telephone: +86 158 3887 2005; fax: 00+86+0379+64830344

Abstract

Obtaining reliable STR genotype of trace DNA is still a difficult problem in forensic genetics. By using different vectors and pretreatment methods to store and handle peripheral blood, this study provides a new idea for obtaining reliable genotype of trace DNA extracted from trace blood samples. FTA cards and medical gauzes gathered blood samples were firstly pretreated by Tris EDTA saline buffer solution (STE) and phosphate buffered saline (PBS). And then DNA was extracted by QIAamp® DNA blood mini kit and genotyped by Human STRtyper-21G amplified fluorescence detection kit on 3130 genetic analyzer. When comparing the performance of pretreatment methods, results showed that STE was better than PBS when FTA cards were used as vector and there was no significant difference between STE and PBS when using medical gauzes as vectors. When comparing vectors, medical gauzes could obtain higher peak height and bigger peak area than FTA card with STE as pretreatment method, but the intra locus balance was poor. Higher peak height and bigger peak area were also found in medical gauze when comparing different vectors with PBS as pretreatment method, while there was no difference in peak height ratio. In conclusion, the vectors and pretreatment methods screened in this research can effectively preserve and dispose the trace peripheral blood samples to perform downstream DNA typing experiments. PBS is recommended for medical gauzes and STE for FTA cards. More carriers, different samples, longer storage time as well as more pretreatment methods are suggested in the future researches.

Keywords
Trace blood sample, Carriers, Pretreatment method, DNA typing.

1. Introduction

Sample collection, storage time or pretreatment method will affect the yield of DNA extraction and successful typing rate of trace DNA [1-4]. Some scholars have pointed out that by using Tris EDTA saline buffer solution (STE) as pretreatment method, the physiological state of dried insect muscle tissue can be restored, which is conducive to the optimization of DNA extraction effect[5-7]. Additionally, pretreatment with phosphate buffered saline (PBS) can significantly improve the DNA extraction effect of alcohol-soaked specimens [5], while the two pretreatment methods' role in human samples remains to be further explored. So, we finally selected FTA card and medical gauze as carriers to determine the impact of different carriers and pretreatment methods on DNA typing efficiency of trace DNA.

2. Materials and methods

2.1. Sample collection

Peripheral blood was placed on FTA cards and medical gauzes, and stored at room temperature for one month. 6 FTA cards and 6 medical gauzes with 5μL peripheral blood in each vector were obtained finally.

2.2. Pretreatment methods

3 FTA cards and 3 medical gauzes were cut and soaked in 200μL STE or PBS buffer solution at room temperature for 24 hours respectively, and then incubated in an electric thermostatic water bath (Shanghai Yuejin Medical Equipment Co., LTD.) at 56°C for 30 minutes.

2.3. DNA extraction and concentration detection

DNA was extracted using the QIAamp® DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol, and the elution volume was set to 20μL. The purity and concentration were determined by FLA6000 microUV-visible spectrophotometer (Jingfei Technology, Hangzhou, China).

2.4. STR typing

Multiple amplification was conducted by STRtyper-21G (HEALTH bioMed, Ningbo, China) which contains 20 autosomal STR (A-STR) and Amelogenin loci. The amplified products were then subjected to capillary electrophoresis on 3130 genetic analyzer (Thermo Fisher Scientific, South San Francisco, CA) and genotyped by GeneMapper™ ID-X v3.2.1 software (Thermo Fisher Scientific).

2.5. Statistical tests

Amelogenin was not considered when performing statistical tests. The alleles with peak height higher than 50 relative fluorescence unit (RFU) were determined as real alleles and the successful typing rates were calculated. The peak height or peak area values of alleles were considered for samples. Peak height ratios (PHR) of heterozygote loci for samples were also analyzed. Locus with PHR between 0.7 (the lower peak heights divide the higher peak heights) and 1.43 (the higher peak heights divide the lower peak heights) were deemed as in a state of peak equilibrium. GraphPad Prism version 7.0 (San Diego, CA, USA) was used for data analysis.

3. Results and Discussion

Alleles of all loci were completely detected and the successful typing rates were 100% for all samples. A total of 20 A-STRs and 35 alleles were analyzed, including 15 heterozygous loci. The results of normality test showed that only the peak height ratio of medical gauzes pretreated by STE was conformed to normal distribution.

When comparing the performance of pretreatment methods, the averages of peak height or peak area for samples pretreated by STE or PBS with FTA cards as vectors were showed that STE was better than PBS when using FTA cards as vectors (Table 1). STE also performed better than PBS when considering PHR. And there was no significant difference between STE and PBS when using medical gauzes as vectors.

When comparing the performance of vectors, significant differences were existed between FTA cards and medical gauzes with STE as pretreatment methods (Table 2). Statistical results addressed that the performance of medical gauzes was better than that of FTA cards. While, it was interested that opposite result was collected when considering PHR which showed that the performance of FTA cards was better than that of medical gauzes. Similar results were found in peak height and peak area when using PBS as pretreatment methods, except there was no significant difference between FTA cards and medical gauzes when considering PHR.

Table 1. The relevant information of samples and statistical tests when comparing pretreatment methods, *p* value which was lower than 0.05 was marked in red.

parameters	vectors	pretreatment methods	averages	*p* value
peak height	FTA card	STE	1995	< 0.01
		PBS	1809	
	medical gauze	STE	3807	> 0.05
		PBS	3781	
peak area	FTA card	STE	27217	< 0.001
		PBS	23638	
	medical gauze	STE	49013	> 0.05
		PBS	47431	
peak height ratio	FTA card	STE	1.11	< 0.05
		PBS	1.21	
	medical gauze	STE	1.20	> 0.05
		PBS	1.17	

Table 2. The relevant information of samples and statistical tests when comparing vectors, *p* value which was lower than 0.05 was marked in red.

parameters	pretreatment methods	vectors	averages	*p* value
peak height	STE	FTA card	1995	< 0.0001
		medical gauze	3807	
	PBS	FTA card	1809	< 0.0001
		medical gauze	3781	
peak area	STE	FTA card	27217	< 0.0001
		medical gauze	49013	
	PBS	FTA card	23638	< 0.0001
		medical gauze	47431	
peak height ratio	STE	FTA card	1.11	< 0.05
		medical gauze	1.20	
	PBS	FTA card	1.20	> 0.05
		medical gauze	1.17	

4. Conclusion

In this study, both the vectors and pretreatment methods can effectively extract trace DNA from microperipheral blood samples stored for one month, and complete STR genotype was obtained from these trace DNA. And the type of vectors needs to be considered when selecting pretreatment method. This study implies that STE is more suitable for the pretreatment of blood samples stored in FTA card, while PBS is more suitable for the pretreatment of blood samples stored in medical gauze. In the future, longer preservation time, multiple carriers as well as more pretreatment methods will be further screened for exploring more suitable strategies to store various types of forensic samples.

5. Acknowledgments

We thank the volunteers in our study. This study was supported by Key scientific research project of Henan University (grant number 24A310002) and Science and Technology Research Projects of Henan Province (grant number 242102310460).

6. Conflict of interest statement

None.

7. References

[1] Xu Q, Wang Z, Kong Q, et al. Evaluating the effects of whole genome amplification strategies for amplifying trace DNA using capillary electrophoresis and massive parallel sequencing [J]. Forensic Sci Int Genet, 2021, 56(102599).

[2] Shujiao D, Ping L I, Guiqing H, et al. Effect of Sample Collection and Preservation on DNA Yield [J]. Forensic Science and Technology, 2015, 40(3): 242-3.

[3] Dillon N, Austin A D, Bartowsky E. Comparison of preservation techniques for DNA extraction from hymenopterous insects [J]. Insect Molecular Biology, 2010, 5(1): 21-4.

[4] Verdon T J, Mitchell R J, Oorschot R A H V. The influence of substrate on DNA transfer and extraction efficiency [J]. Forensic Science International: Genetics, 2013, 7(1): 167-75.

[5] Liu P, Wang W, Yuan C. Comparison of DNA extraction effects of different pretreatment methods on alcohol-impragnated specimens [J]. Journal of Yan'an University (Natural Science Edition), 2022, 41(2): 83-8. (in Chinese)

[6] Huang, Weidong, Xie, et al. Effects of Different Pretreatments of DNA Extraction from Dried Specimens of Ladybird Beetles (Coleoptera: Coccinellidae) [J]. Insects, 2019, 10(4):91-100.

[7] Thongjued K, Chotigeat W, Bumrungsri S, et al. A new cost-effective and fast direct PCR protocol for insects based on PBS buffer [J]. Molecular ecology resources, 2019, 19(3): 691-701.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1073-1078
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Assessment of Diamond™ Nucleic Acid Dye as predictive method for targeting hair follicles suitable for DNA profiling in casework

Leire Palencia-Madrid[1*], Oscar García[1], Roberto Izaga Azkueta[2], José Juan Lata González [2], F. Javier López Curto[2], Juan Antonio Pérez López[1], Eva Romón Martínez[1], Ainara Aldana Rodríguez [1], Estibaliz Cuesta Durán[1], Mónica Encinas Guerra[1], Jokin Ruiz Lopategi[1], Laura Palomo Díaz[1], Mikel González Torres[1]

[1]*Forensic Genetics Section, Forensic Science Unit, Basque Country Police – Ertzaintza; Erandio (Bizkaia), Spain.*

[2]*Forensic Anthropology Section, Forensic Science Unit, Basque Country Police – Ertzaintza; Erandio (Bizkaia), Spain.*

** Corresponding author at: Forensic Science Unit, Ertzaintza. Larrauri Mendotxe Bidea 18, 48950, Erandio (Bizkaia), Spain. leire.palencia@seg.euskadi.eus.*

Abstract

Human hair samples are common in forensic casework, however successful rate of nuclear DNA profiling is low in most of them. Microscopic examination of hair roots reveals the growth stage of the hair and if there are skin cells attached, indicative of the genotyping success. This is a necessary step, since most of the hairs found in crime scenes are in telogen phase, which were naturally shed, and rarely produce informative STR profiles. Staining of hair roots with fluorescent dyes has shown to be a proper method to observe the nuclei present in the sample and predict the possibility to achieve the complete STR genotype of the hair.

Here we present the test performed by our laboratory to assess the use and possible implementation in routine casework of hair staining with Diamond™ Nucleic Acid Dye and fluorescent microscopy observation as screening method to prioritise samples with high probability of genotyping success. After nuclear hair staining and microscopy examination, DNA was extracted, quantified by real-time PCR and genotyped for autosomal and Y chromosome STRs. The method turned out to be quick and easy to use in routine casework processes, and it does not interfere with subsequent DNA analyses. Moreover, hairs that yielded enough amount of DNA to obtain informative STR profiles were clearly identified due to the fluorescence observed at the

proximal hair root end. Thus, it has proved to be a suitable method for implementation in routine casework processes, although further tests are required to improve the results on telogen hairs.

Keywords
Hair, screening, nucleic acid dye, DNA recovery, genotyping success, STR profile.

1. Introduction

In forensic investigations, hair is frequently found at crime scenes [1]. It can either be naturally shed or forcibly removed. Most hair found at crime scenes, up to 90%, is in the telogen phase, meaning it has been naturally shed [2]. However, the genetic material in hairs is often severely degraded, and shed hairs tend to have insufficient nuclear DNA in both quantity and quality, making successful STR typing challenging [3, 4].

Microscopic examination of hair roots reveals the growth stage of the hair and if there are skin cells attached, indicative of the genotyping success [2]. Additionally, staining of hair roots with different type of dyes has shown to be a proper method to observe the nuclei present in the sample, especially for hairs in telogen phase, which have the most limited DNA recovery. The number of observed nuclei predicts the possibility to achieve the complete STR genotype of the hair [5-8].

The aim of this study is to assess the use and possible implementation in our routine casework of the hair samples staining with Diamond™ Nucleic Acid Dye and fluorescent microscopy observation as screening method to prioritise samples with high probability of genotyping success.

2. Materials and methods

Sample collection: shed and plucked head hairs were collected from four donors (two males and two females). Shed and plucked body hairs were also collected from these two males.

Nucleic acid binding dye preparation: a working solution of Diamond™ Nucleic Acid Dye (DD; Promega, Madison, WI, USA) was prepared, diluting 20-fold the stock solution (10,000X) in 75% ethanol.

Nuclear staining and visualisation: 1 µL of working DD solution was applied to the root end of the hair sample placed on a glass slide. They were covered with a coverslip and then visualized under the microscope.

For visualisation, a Leica DM4M microscope (Leica Microsystems, Wetzlar, Germany) with fluorescence module (filter A, UV excitation range, filter excitation 360/40, dichromatic mirror 400, and suppression filter LP 425) was used.

DNA extraction and analysis: DNA was extracted by a standard phenol chloroform extraction procedure, and quantified with the Quantifiler Trio DNA Quantification Kit (Thermofisher Scientific, Waltham, MA, USA; TFS) on a 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA; AB). DNA profiles were generated using up to 1 ng of DNA with the GlobalFiler™ PCR amplification kit (TFS) and the Yfiler Plus™ PCR amplification kit (TFS), on a GeneAmp PCR System 9700 Thermal Cycler (AB). Electrophoresis and STR genotyping were performed in an ABI 3500 Genetic Analyzer (AB) with GeneMapper ID-X v1.6 software (TFS).

3. Results

Plucked hairs showed bright fluorescence at the proximal hair root end (Fig. 1a, 1b). These hairs quantified over 150 pg/µl for the long fragment of Quantifiler Trio DNA kit, and they produced complete genotypes using both GlobalFiler™ (GF) and Yfiler Plus™ (YFP) kits.

Among hairs in anagen phase (Fig. 1a), those that preserved the outer root sheath yielded the maximum DNA, with up to 7 ng/µl (for the long fragment). In these hairs, the fluorescent Diamond™ Nucleic Acid Dye allows us to observe the nuclei of the cells that form the outer root sheath.

On the contrary, a plucked hair in anagen phase that had lost the root end (Fig. 1c), only quantified 11 pg/µl (for the long fragment) and produced partial composite genotypes for GF and YFP kits, with 17 out of 21 loci, and 15 out of 27 loci, respectively.

The fluorescence observed at the root end of shed hairs (Fig. 1d, 1e) is lighter or just residual. These hairs are in telogen phase and they can either keep some remaining root sheath or not. The ones with remnant sheath (Fig. 1d), emitted a slightly more intense fluorescence. These hairs yielded between 7 and 27 pg/µl (for the long fragment), and we could get from them partial composite genotypes for GF (up to 15 out of 21 loci) and YFP kits (up to 20 out of 27 loci). However, in those shed hairs without sheath (Fig. 1e) quantification detected up to 2 pg/µl at the very best, and consequently, the genotyping of these hairs was unsuccessful.

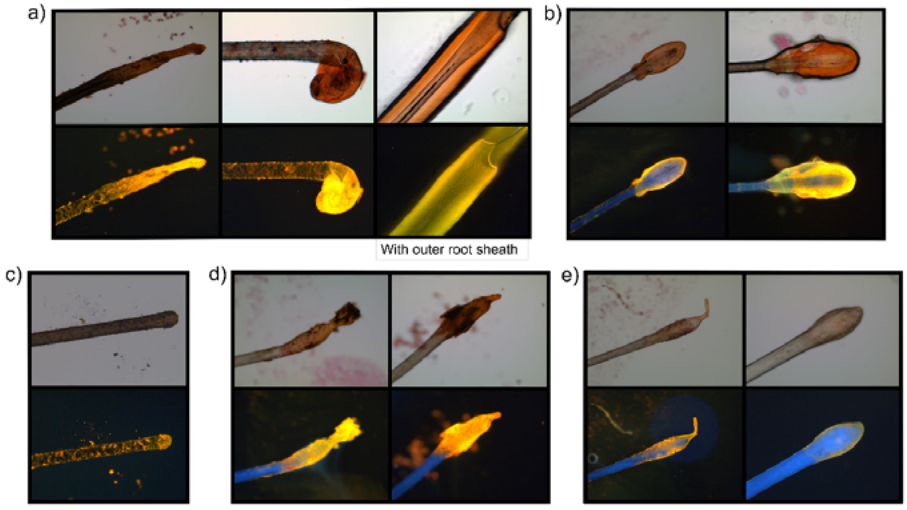


Figure 1. Microscope images at direct and fluorescence visualisation at 100X of the different type of hairs analysed: a) plucked hairs in anagen phase, which produced complete STR genotypes; b) plucked hairs in catagen phase, which produced complete STR genotypes; c) plucked hair in anagen phase (without root), which produced partial STR genotype; d) shed hairs in telogen phase (with remnant sheath), which produced partial STR genotypes; e) shed hairs in telogen phase (without sheath), which produced negative STR genotypes.

4. Discussion

Plucked hairs, which produced complete and correct genotypes, showed bright fluorescence, while a diffuse fluorescence was observed surrounding the root in shed hairs. Moreover, those shed hairs with barely residual or lack of fluorescence produced non-informative STR profiles.

Unlike previous studies, that were able to count cell nuclei at the root of shed hairs [5-8], we only identified stained cell nuclei in those hairs that were accompanied by the outer root sheath (Fig. 1a), that is, plucked hairs. In contrast, despite the residual fluorescence, no nuclei could be differentiated in shed hairs. This issue may be due to the microscope's filter configuration (Alicia M. Haines, personal communication). By this reason, and with the aim of visualising the nuclei present on telogen hair roots, further adjustments would be done to improve and tune up the technique's performance.

5. Conclusion

From these results, several insights can be gleaned. The use of DD for microscopic examination of hair samples does not interfere with subsequent DNA extraction and PCR analyses. Plucked hairs in anagen and catagen phases yield enough DNA to produce complete STR profiles. Anagen hairs without root still can produce partial STR profiles. Between shed hairs, telogen hairs that preserve remnant root sheath can also produce partial STR profiles, while in those without sheath remains the DNA recovery is insufficient for STR genotyping. Therefore, DD can be especially helpful on telogen hairs, to clarify whether they bring some remnant root sheath and are able to produce partial STR profiles or not.

6. Acknowledgments

This research received no external funding.

7. Conflict of interest statement

The authors declare no conflict of interest.

8. References

[1] C.T. Oien. Forensic hair comparison: background information for interpretation. *Forensic Science Communications*. 2009;11(2).

[2] E.A. Graham. DNA reviews: hair. *Forensic Sci Med Pathol*. 2007;3(2):133-137. doi.org/10.1007/s12024-007-9005-9

[3] C.A. Linch. Degeneration of nuclei and mitochondria in human hairs. *J Forensic Sci*. 2009;54(2):346-349. doi.org/10.1111/j.1556-4029.2008.00972.x

[4] M.D. Brandhagen, O. Loreille, J.A. Irwin. Fragmented Nuclear DNA Is the Predominant Genetic Material in Human Hair Shafts. *Genes*. 2018; 9(12):640. https://doi.org/10.3390/genes9120640

[5] L. Bourguignon, B. Hoste, T. Boonen, *et al*. A fluorescent microscopy-screening test for efficient STR-typing of telogen hair roots. *Forensic Sci Int Genet*. 2008;3(1):27-31. doi.org/10.1016/j.fsigen.2008.08.006

[6] T. Lepez, M. Vandewoestyne, D. Van Hoofstat, *et al*. Fast nuclear staining of head hair roots as a screening method for successful STR analysis in forensics. *Forensic Sci Int Genet*. 2014;13:191-194. doi.org/10.1016/j.fsigen.2014.08.003

[7] A.M. Haines and A. Linacre. A rapid screening method using DNA binding dyes to determine whether hair follicles have sufficient DNA for successful profiling. *Forensic Sci Int*. 2016;262:190-195. doi.org/10.1016/j.forsciint.2016.03.026

[8] S.Y. Lee, E.J. Ha, S.K. Woo, *et al*. A rapid nuclear staining test using cationic dyes contributes to efficient STR analysis of telogen hair roots. *Electrophoresis*. 2017;38(13-14):1771-1779. doi.org/10.1002/elps.201700024

9. Figures

Figure 1. Microscope images at direct and fluorescence visualisation at 100X of the different type of hairs analysed: a) plucked hairs in anagen phase, which produced complete STR genotypes; b) plucked hairs in catagen phase, which produced complete STR genotypes; c) plucked hair in anagen phase (without root), which produced partial STR genotype; d) shed hairs in telogen phase (with remnant sheath), which produced partial STR genotypes; e) shed hairs in telogen phase (without sheath), which produced negative STR genotypes.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1079-1084
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Deletion events at DYS448 locus: Exploring rearrangement patterns in the AZFc region of the Y chromosome

R. Flores-Espinoza[1*], M. Nguidi[1], J. Ribeiro[1], M. Calvopiña[2], G. Burgos[2, 3], L. Gusmão[1]

[1] *Laboratório de Diagnóstico por DNA (LDD), Universidade do Estado do Rio de Janeiro, Brazil*

[2] *One Health Research Group, Facultad de Medicina, Universidad de las Américas, Quito, Ecuador*

[3] *Grupo de Medicina Xenómica, Instituto de Ciencias Forenses, Universidad de Santiago de Compostela, Spain*

**Corresponding author at: Laboratório de Diagnóstico por DNA (LDD), Universidade do Estado do Rio de Janeiro, Brazil., 20550-900. Email address: rodrigofabian.ffe@gmail.com*

Abstract

The ampliconic sequence of the human Y chromosome contains numerous palindromes and inverted repeats, making it susceptible to non-allelic homologous recombination (NAHR) events, inducing chromosome rearrangements. Here, we report and discuss the causes of rearrangements within the AZFc region, associated with the deletion of the DYS448 locus. Unexpected rearrangements were observed, which were explained by non-NAHR-mediated events. Due to the low frequency of non-NAHR events, associations with specific haplogroups are expected to be stronger than those for NAHR rearrangements.

Keywords

Y-STR, Null allele, NAHR, STS markers

1. Introduction

The ampliconic sequences of the human Y chromosome contain abundant palindromes and inverted repeats that are prone to non-allelic homologous recombination (NAHR), inducing deletions or duplications [1, 2]. These sequences are located mainly on the long arm of the Y chromosome, in the three domains of the azoospermia factor regions AZFa, AZFb and AZFc. The AZFc is predominantly composed

of duplicated sequences known as amplicons, which exhibit high sequence identity. Due to this high sequence identity, NAHR frequently occurs between the amplicons, potentially leading to chromosome rearrangements [3]. Analyses with different sets of Y-STRs revealed null alleles at DYS448, due either to a complete deletion of the locus or point mutations at primer binding sites. Accordingly, two main AZFc partial deletions involving DYS448 have been identified, resulting from NAHRs between the palindromic segments b1/b2 and b1/b3 [4–8]. Although less frequently, alternative mechanisms not mediated by NAHR have been also described [9]. In this study, we describe the rearrangements within the AZFc region that were responsible for null alleles at DYS448, and their association with specific haplogroups.

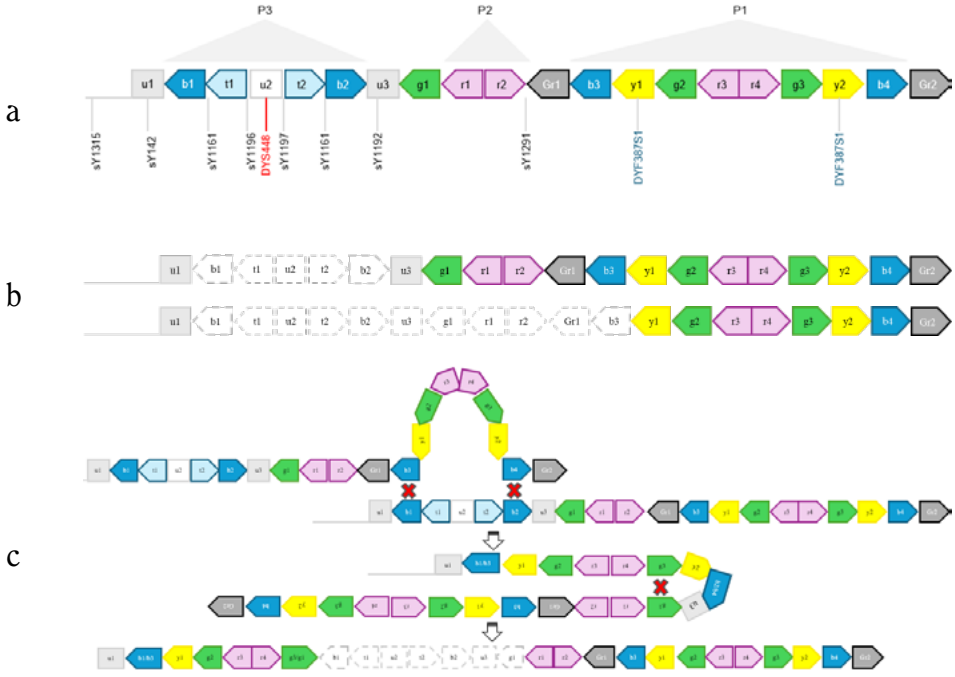
2. Materials and Methods

DNA samples from nine males from Ecuador, Brazil and Colombia were genotyped for three different Y-chromosomal marker sets: (i) 23 Y-STRs included in the PowerPlex® Y-23 System (Promega Corporation); (ii) 27 Y-STRs included in the Yfiler™ Plus System (Applied Biosystems™); and 28 Y-STRs included in the Investigator Argus Y-28 QS (QIAGEN). Amplification was performed according to the manufacturer's instructions for each kit. Haplogroups were predicted using NEVGEN (https://www.nevgen.org/) based on the Y-STR haplotype data. All samples were screened using Y-specific sequence-tagged sites (STSs) covering the Yq11.223 zone. The STSs, and corresponding PCR primers, were selected using the UCSC Genome Browser hg19 assembly (GRCh37/hg19) (http://www.genome.ucsc.edu/). The STSs were assembled in two multiplex reactions (M1: sY142, sY1315, sY1196, sY1291; M2: sY1192, sY1161, sY1197). PCR was performed in 10 µL final volume reaction, using 1x Qiagen multiplex PCR master mix (QIAGEN), 2 ng of DNA, and 0.2 mM of each primer. The amplification consisted of a denaturing step at 95 °C for 15 minutes, followed by 35 cycles of denaturation at 94 °C for 30 seconds, annealing at 60 °C for 90 seconds, extension at 72 °C for 60 seconds, and a final extension step at 72 °C for 30 minutes. Separation and detection of the amplicons were performed by electrophoresis in a 3% agarose gel.

3. Results

Five different rearrangements were identified, four of them explained by NAHR events that led to AZFc partial deletions. Two out of the nine samples showed an intrachromatidic b1/b2 deletion, and four have an interchromatidic b1/b3 deletion

(Fig. 1b). Another sample evidenced a partial deletion (Fig. 1c), explained by two simultaneous and one subsequent NAHRs: (a) b1/b3 deletion and b2/b4 duplication; and (b) an intrachromatidic g3/g1 deletion. This AZFc partial deletion cannot be explained by two independent and one subsequent event. Namely, if the deletion occurred first, subsequent events such as b2/b4 duplication or g3/g1 deletion would not be possible in the resulting chromosome. In one case, a discontinued partial deletion was detected (Fig. 1d), being explained by two subsequent NAHR-mediated intrachromatidic deletions: t1/t2 and g3/g1. In another case, the absence of the flanking markers sY142 and sY1315 (Fig. 1e) suggests that the deletion was not caused by NAHR or conversion between amplicons. In this case, a sporadic event is more likely to have occurred.



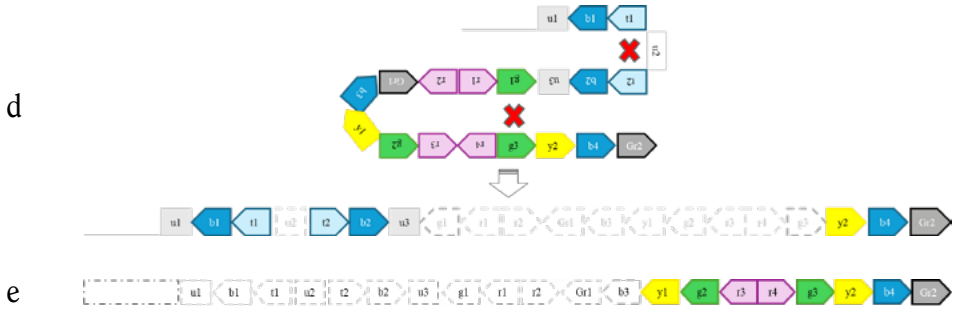


Figure 1. (a) Y chromosome structure, showing the location of the AZFc region and the Y-STS markers used. Multiple copies of six amplicons are represented by color-coded arrows (b-blue, t-sky-blue, g-green, r-pink, y-yellow, and Gr-grey). The arrow's direction indicates amplicon copy orientation. The white rectangle between t1 and t2 represents the large P3 spacer, while the grey rectangles represent a single copy of inverted repeats. The dashed grey lines represent the deleted regions.

4. Discussion

A network was constructed for our samples, using the Y-STR data (Fig. 2). The five chromosomes with a b1/b3 deletion belong to two sub-haplogroups: R1a-M198 (S1-S4) and R1b-M269 (S5). A chromosome belonging to haplogroup E1b1a showed an NAHR event (Fig. 1c), which was previously described in a sample from haplogroup R1b-V69 [6] (S9 in Fig. 2). These findings, involving the same deletion in different haplogroup backgrounds, provide evidence of independent mutations via NAHR mechanisms. In contrast, two chromosomes with a b1/b2 deletion (S6 and S8 in Fig. 2) belonging to the same haplogroup, suggest a common origin within haplogroup Q. Additionally, a sample showing a non-NAHR event (Fig. 1e; S7 in Fig. 2) also belongs to haplogroup Q. The low frequency of non-NAHR events suggests a possible association of this rearrangement with haplogroup Q. Supporting this hypothesis is the high frequency of null alleles at the DYS448 locus in Native American populations. A search performed in the Y Chromosome Haplotype Reference Database (YHRD, www.yhrd.org) showed high frequencies of null alleles at this locus (varying between 0.001 in Europeans and 0.01 in Native Americans), being at least five times more frequent in Native Americans compared to other metapopulations (Africa, Europe, East Asia, and Admixed American).

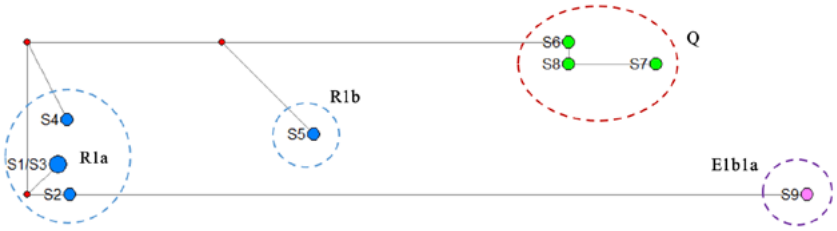


Figure 2. Network representing nine individuals carrying Y chromosome rearrangements. Blue, green, and pink dots indicate Mestizos, Native Americans, and Afro-descendants, respectively. Red dots represent the median nodes of the network, and the dot area is proportional to the haplotype frequency.

5. Conclusion

In this study, five different rearrangements were characterized. Four were explained by NAHR, while one was attributed to a non-NAHR mechanism. The low frequency of non-NAHR events and the high frequency of null alleles at DYS448 in Native American populations suggest a possible association between this unexpected rearrangement and haplogroup Q.

6. Acknowledgements

RF and LG were supported by Fundação de Amparo à Pesquisa do Rio de Janeiro – FAPERJ, (E-26/202.525/2024, CNE-2022 and E-26/211.369/2021), Brazil. LG was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (ref. 315007/2023-0), Brazil. MN was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES). GB was supported by MED.GBF.22.03 grant from DGIV, Universidad de Las Américas, Quito, Ecuador.

7. Conflict of interest statement

None.

8. References

[1] Trombetta B, Cruciani F. Y chromosome palindromes and gene conversion. *Hum Genet* 2017; 136: 605–619.

[2] Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 2003; 423: 825–837.

[3] Kuroda-Kawaguchi T, Skaletsky H, Brown LG, et al. The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat Genet* 2001; 29: 279–286.

[4] Hajiesmaeil M, Ravasini F, Risi F, et al. High incidence of AZF duplications in clan-structured Iranian populations detected through Y chromosome sequencing read depth analysis. *Sci Rep*; 13. Epub ahead of print 1 December 2023. DOI: 10.1038/s41598-023-39069-7.

[5] Ravasini F, D'Atanasio E, Bonito M, et al. Sequence Read Depth Analysis of a Monophyletic Cluster of Y Chromosomes Characterized by Structural Rearrangements in the AZFc Region Resulting in DYS448 Deletion and DYF387S1 Duplication. *Front Genet*; 12. Epub ahead of print 16 April 2021. DOI: 10.3389/fgene.2021.669405.

[6] Turrina S, Caratti S, Ferrian M, et al. Deletion and duplication at DYS448 and DYS626 loci: unexpected patterns within the AZFc region of the Y-chromosome. *International Journal of Legal Medicine* 2015; 129: 449–455.

[7] Repping S, Skaletsky H, Brown L, et al. Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nat Genet* 2003; 35: 247–251.

[8] Liu X, Li Z, Su Z, et al. Novel Y-chromosomal microdeletions associated with non-obstructive azoospermia uncovered by high throughput sequencing of sequence-tagged sites (STSs). *Sci Rep* 2016; 6: 21831.

[9] Balaresque P, Bowden GR, Parkin EJ, et al. Dynamic nature of the proximal AZFc region of the human Y chromosome: Multiple independent deletion and duplication events revealed by microsatellite analysis. *Hum Mutat* 2008; 29: 1171–1180.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1085-1090
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Investigation of a Y-chromosome deletion affecting amelogenin-based sex typing in forensics using CE and MPS methods

Magdalena Marcińska[1], Agnieszka Parys-Proszek[1], Maria Wróbel[1]

[1] *Institute of Forensic Research, Kraków, Poland.*

** Magdalena Marcińska at: Westerplatte 9, 30-033 Kraków, mmarcinska@ies.gov.pl,*

Abstract

Sex determination in forensic science is particularly important in human identification cases and sexual offence crimes. The amelogenin marker (AMEL), which enables sex typing, was described in the early 1990s and immediately found widespread use in human identification studies. Its analysis is based on standard capillary electrophoresis (CE) and can be combined with analysis of STR markers. The presence of the Y chromosome (ChrY) amelogenin variant in the genetic profile clearly indicates male sex. However, cases of lack of detection of the ChrY allele, known as AMELY-null, have been reported. This can lead to misidentification of the female sex, resulting in misleading information being reported. The aim of the project was a thorough investigation of the AMELY-null case observed in a forensic sample, using CE and massively parallel sequencing (MPS) methods. DNA profiling using NGM and GlobalFiler kits showed lack of the Y allele within the amelogenin marker. Analysis of ChrY STR markers (Y-filer Plus and PowerPlex Y23 kits) revealed deletion of 6 markers closely located on the short arm of ChrY. MPS using ForenSeq kit confirmed the occurrence of mutation in the tested sample. Further interpretation of the obtained results revealed a large deletion encompassing 7 loci (DYS570, DYS576, DYS522, DYS458, DYS449, DYS481, DYS627) in the analyzed sample, which is consistent with the common observation of susceptibility of certain regions of the Y chromosome to structural mutation events. The use of various Y-STR multiplex kits containing different marker primers and different DNA analysis techniques proves that the presence of the AMELY-null allele is not due to primer binding site mutations. Our study confirmed that AMELY sex typing can be misleading and the female sex should be interpreted in the light of other markers useful in sex typing including ChrY-STRs, ChrY-indels, and those analysed at the stage of evaluation of DNA concentration with e.g., Quantifiler Trio DNA Quantification method. The occurrence of ChrY deletions involving amelogenin locus is rare but particular caution is recommended in sex typing for populations more susceptible to AMELY deletions (e.g. South Asian region).

Keywords
AMELY deletion, Sex typing, STR marker

1. Introduction

Sex typing is one of the important components of individual identification in criminal investigations. In humans, sex-typing genes known as Amelogenins (AMEL) are presented on both the X and the Y chromosomes. The amelogenin marker was described in the early 1990s and immediately found widespread use as a routine tool for sex typing in forensic DNA analysis. Analysis of this sex marker is based on the detection of length variations in the X and Y homologous amelogenin gene (AMELX and AMELY) and is performed in conjunction with analysis of STR markers using standard capillary electrophoresis (CE). The presence of the Y chromosome (ChrY) amelogenin variant in the genetic profile clearly indicates male sex. However, cases of lack of detection of the ChrY allele, known as AMELY-null, have been reported. This can lead to misidentification of the female sex implicating serious consequences in forensic inference particularly in human identification cases and sexual offence crimes [1]. The aim of the project was a thorough investigation of the AMELY-null case observed in a forensic sample, using CE and massively parallel sequencing (MPS) methods.

2. Material studied, methods, techniques

The sample for the study was taken from the routine casework as a cigarette butt from a burglary case. There was no reference material. The part of the cigarette butt was subjected to DNA extraction with QIAsymphony DNA Investigator Kit (*Qiagen*) on Qiasymphony SP/AS system. Quantification of sample was performed by using Quantifiler Trio DNA Quantification Kit *(ThermoFisher Scientific)*. Amplification of 1 ng quantified DNA was done using NGM and GlobalFiler system kits *(ThermoFisher Scientific)* including autosomal STRs. Amplification of Y specific markers was carried out by using Y-filer Plus *(ThermoFisher Scientific)* and PowerPlex Y23 *(Promega)* system kits. PCR products were analysed with capillary electrophoresis. Apart from CE technology, an analysis using MPS technology on Miseq FGx (Illumina) sequencing platform with ForenSeq DNA Signature Prep kit *(Verogen)* was applied.

3. Results

DNA profiling using NGM kit showed lack of the Y allele within the amelogenin marker despite the presence of male DNA reported in the Y-target DNA quantification of a sample (1,9 ng/µl). The result was confirmed with the analysis using GlobalFiler kit (Figure 1). Observed missing Y allele in amelogenin should be considered as a dropout of Y-specific amelogenin because of successful amplification of two additional Y specific markers including DYS391 and Yindel incorporated in the multiplexing system. This observation was further reinforced by amplification of 25 loci of Y-chromosome with Y-filerPlus system kit, which revealed deletion of 6 markers (Figure 2). These deletions were also observed within common loci of PowerPlex Y23 kit. The occurrence of mutations detected by CE methods was confirmed with MPS methods using ForenSeq kit system, which additionally revealed lack of DYS522.

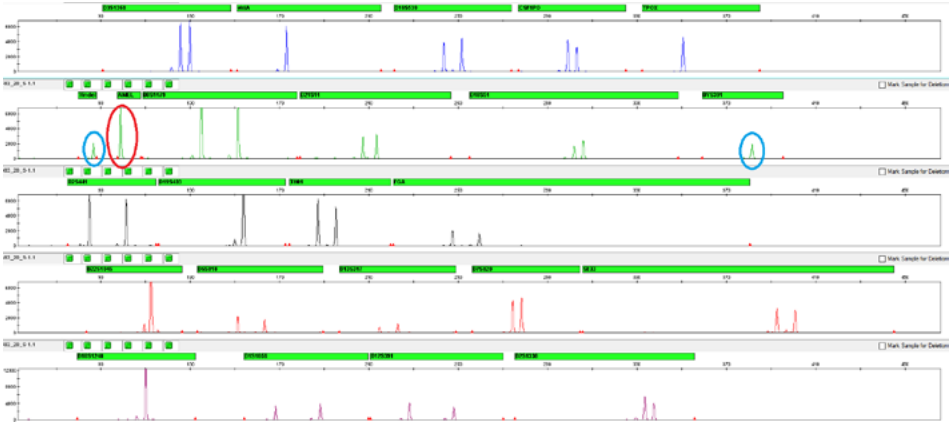


Figure 1. Electropherogram showing AMELY dropout (marked in red) and amplification of DYS391 and Yindel (marked in blue) in a sample analysed with GlobalFiler kit.

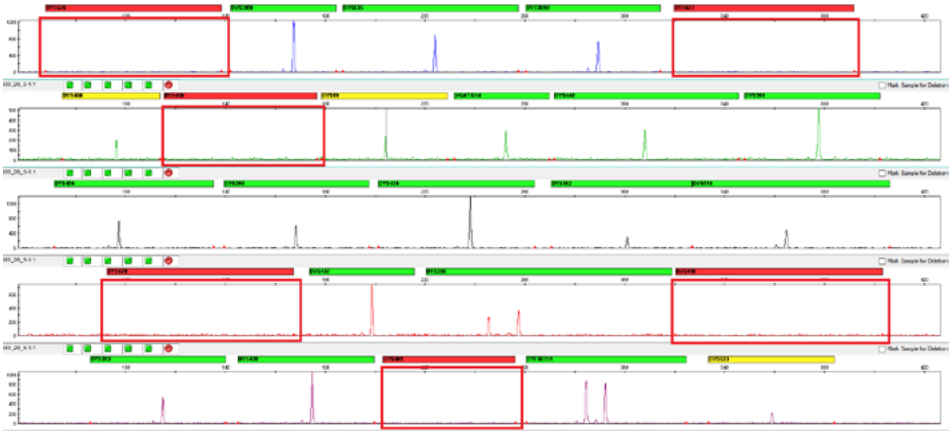


Figure 2. Electropherogram showing deletion at 6 loci (marked in red) in a sample analysed with Y-filerPlus kit.

4. Discussion

Comprehensive interpretation of the results revealed a large size of the deletion encompassing 7 loci (DYS570, DYS576, DYS522, DYS458, DYS449, DYS481, DYS627) closely located on the short arm of ChrY (Yp11.2) (Figure 3). This observation is consistent with the common tendency of susceptibility of certain regions of the Y chromosome to structural mutation events [2].

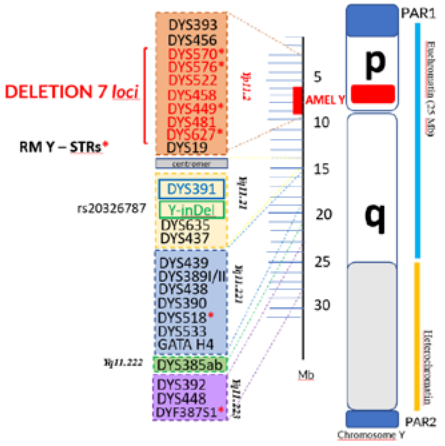


Figure 3. The deletion map of 7 loci marked in red within Yp11.2 fragment with approximate locations of markers [based on 2].

5. Conclusion

In this study we report an investigation of sample with deletion of fragment Y-chromosome among amelogenin marker. The use of various Y-STR multiplex kits containing different marker primers and incorporating different DNA analysis techniques proves that the presence of the AMELY-null allele is not due to primer binding site mutations, which indicates that it represents a case of structural variation. Our study confirmed that AMELY sex typing can be misleading and the female sex should be interpreted in the light of additional markers useful in sex typing including ChrY-STRs, ChrY-indels, and those analysed at the stage of evaluation of DNA concentration with e.g., Quantifiler Trio DNA Quantification method [3]. It is well known that there are ethnic differences in the frequency of AMELY deletion, fluctuating from 0.018% to 8% [4]. Although AMELY null males are extremely rare in most populations, particular caution is recommended in drawing conclusions for individuals originating from populations more susceptible to AMELY deletions coming from India, Nepal, Sri Lanka [5].

6. Acknowledgments

The study was financially supported from the resources of the Institute of Forensic Research in Krakow, grant no. II/G/2022-2023.

7. Conflict of interest statement

None.

8. References

[1] A. Kumar. Implications of Male Amelogenin Dropouts in Forensics. Journal of Forensic Sciences, vol. 15, no. 2, 2021.

[2] D. Syndercombe Court. The Y chromosome and its use in forensic DNA. Emerging Topics in Life Sciences (2021).

[3] N. Kumarm A. Chauhan, R. Gupta, SK Shukla. Role of 3-Y markers in Y deleted allele on amelogenin in autosomal STR. International Journal of Molecular Biology, 4(2):75-78, 2019.

[4] E. Butler, R. Li. Genetic markers for sex identification in forensic DNA analysis. Journal Forensic Investigation, vol. 2, no. 3, pp.1-10, 2014.

[5] K. Masuyama,H. Shojo, H. Nakanishi, S. Inokuchi, N. Adachi. Sex Determination from Fragmented and Degenerated DNA by Amplified Product-Length Polymorphism Bidirectional SNP A nalysis of Amelogenin and SRY Genes. PLOS One (2017).

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1091-1098
DOI: https://dx.doi.org/10.15304/cc.2025.1869

A practical proposal for interpreting X-STR profiles from the Forenseq DNA Signature Prep

Irán Cortés-Trujillo[1], Eduardo Rojas-Prado[1], Mayra Elizabeth García-Aceves[1], Guadalupe Avalos-Navarro[1], Gabriela Martínez-Cortés[1], José Alonso Aguilar-Velázquez[2], Héctor Rangel-Villalobos[1*]

[1]*Instituto de Investigación en Genética Molecular, Departamento de Ciencias Medicas y de la Vida, Centro Universitario de la Ciénega, Universidad de Guadalajara (CUCI-UdeG), CP 47810, Ocotlán, Jalisco, México.*

[2]*Instituto de Investigación en Ciencias Biomédicas, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara, Jalisco, México*

** Corresponding Author: Dr. Héctor Rangel Villalobos*
Instituto de Investigación en Genética Molecular, Universidad de Guadalajara.
Centro Universitario de la Ciénega (CUCiénega-UdeG)
Av. Universidad #1115, Col. Paso Blanco, CP 47810 · Ocotlán, Jalisco, México
Tel. +52 (392)1000979; fax +52 (392)9257112
E-mail: hrangel13@hotmail.com

Abstract

The ForenSeq DNA Signature Prep kit (Verogen) analyzed by Massively parallel sequencing (MPS) includes 7 X-STRs, which work as a complementary tool to disclose complex kinship cases. However, X-STR databases of Repeat sequence-based (RSB) alleles from MPS, are not currently representative of most human populations. Interestingly, all the 7 X-STRs of the ForenSeq kit are a subset of the ArgusX-12 kit (Qiagen), the X-STR kit most studied worldwide. In addition, because of the compatibility between RSB and Length-based (LB) alleles, the existing large population databases of LB alleles can be useful for forensic casework interpretation of X-STR profiles from the ForenSeq kit. For practical purposes, those 7 X-STR population databases should be in the FamlinkX format, the software employed for kinship estimation with X-STRs. In this work, we assessed this possibility with an updated Mexican database for ArgusX12 (from 933 to 1155) (https://www.famlink.se/fx_download.html). In four complex kinship cases, we estimated three likelihood ratios (LRs) considering three population database options with the FamlinkX software: *a)* Original ArgusX12 (n= 933); *b)* Updated ArgusX-12 (n= 1155); *c)* 7 X-STRs of the ForenSeq (n= 1155). Interestingly, the updated Mexican population database (n= 1155)

increased significantly the LRs estimated from these kinship cases (average of increase 183.1%). Conversely, LRs from 7 X-STRs represented -in average- only 0.8% of the values provided by 12 X-STRs (*b vs. c*). In brief, we evaluated the forensic impact of updating the sample size of one X-STR population database, and we showed a landscape of the *a posteriori* offered by different X-STR systems (12 vs 7 loci),

Keywords
MPS; Argus X-12; FamlinkX; X-STRs; Mexican population.

1. Introduction

Short tandem repeat loci on the X chromosome (X-STRs) have been integrated into the battery of available markers in forensic genetic labs because their peculiar inheritance pattern helps to solve complex kinship cases [1-3]. For biostatistical interpretation of X-STR profiles, the FamLinkX software takes into account X-linked inheritance peculiarities to estimate the likelihood ratio (LR) in different complex kinship cases [4-5]. The Investigator Argus X-12 system (Qiagen) analyzes 12 X-STRs clustered into four linkage groups, and many worldwide population datasets compatible with FamlinkX are available for interpretation purposes (http://www.famlink.se/fx_download.html). However, considering the large number of possible haplotypes for each linkage group, most of the available worldwide population datasets for FamlinkX underrepresent the actual haplotype diversity [6].

The Massivelly parallel sequencing (MPS) technology has been included in some forensic genetic labs to solve complex cases. Among the forensic genomic platforms based on MPS, the ForenSeq™ DNA Signature Prep Kit (Verogen) combines the simultaneous analysis of STRs with different inheritance patterns, plus SNPs [7]. These genomic systems allow the detection STR alleles with the same length but different sequences, which is known as repeat sequence-based (RSB) variation, besides the Length-based (LB) variation offered by the standard capillary electrophoresis (CE) analysis [8]. Interestingly, the ForenSeq kit analyzes 7 of the 12 X-STRs included in the Argus X-12 kit, including three two-loci linkage groups (LGs): DXS10135-DXS8378 (LG1), DXS7131-DXS10074 (LG2), DXS10103-HPRT (LG3), plus DXS7423 of the LG4 [7] (Figure 1). Unfortunately, few RSB alleles and haplotypes for these 7 X-STRs have been described in the first worldwide population studies with ForenSeq, such as those reported in Mexico (n= 45) [9]. Because of the expected compatibility between RSB and LB alleles, it is possible -and far cheaper- to create 7 X-STR population databases from the available LB alleles of the

Argus X-12 kit [10], than from RSB alleles with the ForenSeq kit [9]. Although lower allele diversity is expected in LB than RSB alleles, the more robust population sample size of LB databases will offer better kinship estimates for interpretation purposes.

In this work, we assessed this practical option for evaluating four kinship cases with FamlinkX and using an updated Mexican database for ArgusX12 (from 933 to 1155) [10]. We demonstrated how LB allele population databases constitute a practical option for the interpretation of complex cases with X-STRs of the Forenseq kit, and the impact of updating the population sample size.

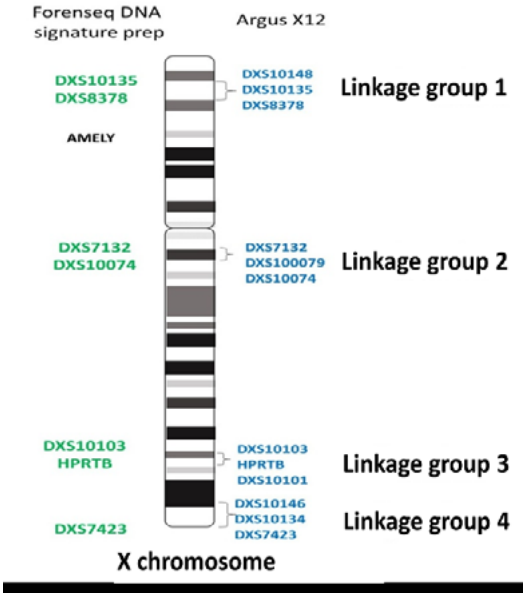


Figure 1. Distribution of X-STRs and linkage groups of the Forenseq Kit with respect to the Argus x-12 kit

2. Methods

We collected blood stain samples impregnated on FTA papers from 74 families including the father, mother, and daughter. The paternity was previously confirmed with the Powerplex Fusion system (Promega Corp.). Bloodstain punches of 1.2 mm washed with FTA Purification Reagent (Whatman™) were used as DNA samples for

genotyping. Participants signed an informed consent letter approved by the Ethical Research Committee at the Instituto de Investigación de Genética Molecular of the CUCiénega, University of Guadalajara (CUCienega-UdeG). This study follows the regulations of the General Health Law on research involving human subjects (Mexico) and the Helsinki Declaration. The anonymity of the volunteers was preserved at all times.

The PCR amplification was carried out with the Investigator Argus X-12 QS kit (Qiagen, Hilden, Germany), followed by capillary electrophoresis in an ABI-Prism 3130 genetic analyzer (Applied Biosystems®, Foster City, USA) according to the manufacturer recommendations. For allele calling, we used the Genemapper version 3.2 software (Applied Biosystems® Foster City, USA) helped by the corresponding allelic ladder. Our laboratory participates with the Argus X-12 QS system in the quality control proficiency test annually organized by the Spanish and Portuguese-Speaking Working Group of the International Society for Forensic Genetics (www.ghep-isfg.org). A total of 222 X-STR haplotypes for ArgusX-12 were obtained from the 74 families by subtracting the paternal haplotypes from the daughter genotype, which allowed discrimination of maternal haplotypes.

3. Results and Discussion

We obtained 222 haplotypes for the ArgusX-12 kit to update the Mexican population database, from 933 to 1155 haplotypes [10]. A total of 234 alleles were observed, with seven additional alleles than the previous Mexican study (n= 933) [10], but with similar modal allele patterns and informativity parameters (data not shown).

In the updated Mexican population sample of Argus X-12, 1508 different haplotypes were observed among the four linkage groups. This means -in average- 243 (19.2%) more haplotypes by linkage group (LG) than in the previous Mexican database [10]. The haplotype number increment ranged from 11.6% for the LG1 (n= 45), to 25.1% for the LG2 (n= 75). Based on the allele number observed by locus, the number of possible haplotypes among LGs in our Mexican population sample ranged from 1638 (LG3) to 9207 (LG4) for Argus X-12.

In addition, we generated a subset of the ArgusX-12 database for the 7 X-STRs of the ForenSeq kit. The possible number of haplotypes for these 7-X-STRs diminished drastically from 224 (LG1) to 63 (LG3) possible haplotypes among three linkage groups, whereas DXS7423 (in the LG4 of ArgusX-12) provides only nine alleles to complete the haplotype diversity of the Forenseq kit. Between these two X-STR

systems, the number of haplotypes for Argus X-12 is theoretically 13 million higher than those of the Forenseq (4.22 E+14 vs 3.2 E+07).

Two X-STR population databases compatible with the FamlinkX software were created [5]. The first one comprises the updated Mexican population database of ArgusX-12 that better represents the haplotype diversity [6], and to get more confident likelihood ratios (LRs). The second one involves the LB database for interpreting complex kinship cases analyzed with the 7 X-STRs of the ForenSeq kit, probably this is the first one reported worldwide for this purpose (available upon request). These FamlinkX files will be helpful in Mexico -and Latin America- to make the forensic interpretation with X-STRs, given the scarce number of worldwide population databases.

We assessed the impact of the population size update done to the Argus X-12 database. We compared the exact LR from four complex kinship cases analyzed with the available FamlinkX file from Mexico (n= 933) [10], and the updated one described herein (n= 1155) (Figure 2). In all these kinship cases, the exact LR from the updated database increased from 135 to 262% concerning the reported database (average LR increment = 183.1%). In addition, we estimated the LR of the same kinship cases with 7 X-STRs. The LR values with 7 X-STRs represented -on average- 0.8% of those supplied by 12 X-STRs (range: 0.01% –1.22%), with the updated population size database (Figure 2). Interestingly, assuming that LR ≥ 1000 is an acceptable kinship level (W= 99.9%), with Argus X-12 only the maternal half-sister case would be below this level. On the contrary, with 7 X-STRs, only the paternal half-sister case with data mothers reached conclusive results (Figure 2).

In brief, our results suggest that the 7 X-STRs of the Forenseq kit offer limited a posteriori informativity to solve most complex kinship cases. Although the larger variability of RSB alleles for STRs obtained by MPS probably will help to rectify this situation [9], the sample size of the RSB population databases presently impacts the forensic casework interpretation, as also is demonstrated herein (Figure 2). Efforts to create X-STR metapopulation databases, for instance at the continental level, probably will make it easier to rectify the population sample size limitations for forensic casework interpretation purposes.

EXACT LR Complex kinship cases X-STR system (sample size)	HALF SISTER CASES			Paternal grandmother-granddaughter
	Maternal 	Paternal 	Paternal (Data mothers) 	
1) Argus X12 (n=933)	111.45	5,566.89	6.034E+07	15,864.8
2) Argus X12 update (n=1155)	292.33	9,324.67	1.01E+08	21,469.3
3) 7 X-STRs ForenSeq (n=1155)	3.56	153.46	12,270.8	89.30
Exact LR increase by population sample size: *2) respect to 1)*	**262.3 %**	**167.5 %**	**167.4 %**	**135.3 %**
Exact LR decrease by loci number *3) respect to 2)*	**1.22 %**	**1.65 %**	**0.01 %**	**0.42 %**

Figure 2. Comparison between exact LRs generated by FamilinkX from the same four complex kinship cases with different Mexican population databases: 1) Original database (n= 933) [10]; 2) Updated database (n= 1155); 3) 7 X-STRs (n= 1155).

4. Conclusion

We updated the Mexican population database compatible with the FamlinkX software for the Argus X12 kit. We created the first LB allele database useful for interpreting X-STR profiles from the ForenSeq kit. We confirmed the impact of the sample size in X-STR population databases for the interpretation of complex kinship cases.

5. Acknowledgments

We thank Consejo Nacional de Ciencia y Tecnología CONACyT (Mexico), for the research grant to H.R-V (N°286623), and for a scholarship to E.R-P (N° 17535). QIAGEN (Mexico) for support with reagents to H.R-V (without number).

6. Conflict of interest statement

The authors declare no competing interests.

7. References

[1] Diegoli TM (2015) Forensic typing of short tandem repeat markers on the X and Y chromosomes. Forensic Sci Int Genet 18:140-151. https://doi.org/10.1016/j.fsigen.2015.03.013

[2] Pinto N, Gusmao L, Amorim A (2011) X-chromosome markers in kinship testing: a generalization of the IBD approach identifying situations where their contribution is crucial, Forensic Sci Int Genet 5:27-32 https://doi.org/10.1016/j.fsigen.2010.01.011

[3] Tillmar AO, Kling D, Butler JM, Parson W, Prinz M, Schneider PM, Egeland T, Gusmão L (2017) DNA Commission of the International Society for Forensic Genetics (ISFG): Guidelines on the use of X-STRs in kinship analysis, Forensic Sci Int Genet 29:269-275 https://doi.org/10.1016/j.fsigen.2017.05.005

[4] Tillmar AO, Egeland T, Lindblom B, Holmlund G, Mostad P (2010) Using X-chromosomal markers in relationship testing: calculation of likelihood ratios taking both linkage and linkage disequilibrium into account. Forensic Sci Int Genet 5:506-511 https://doi.org/10.1016/j.fsigen.2010.11.004

[5] Kling D, Dell'Amico B, Tillmar AO (2015) FamLinkX–implementation of a general model for likelihood computations for X-chromosomal marker data, Forensic Sci Int Genet 17:1-7 https://doi.org/10.1016/j.fsigen.2015.02.007

[6] Pinto N, Pereira V, Tomas C, Loiola S, Carvalho EF, Modesti N, Maxzud M, Marcucci V, Cano H, Cicarelli R, Januario B, Bento A, Brito P, Burgos G, Paz-Cruz E, Díez-Juárez L, Vannelli S, Pontes ML, Berardi G, Furfuro S, Fernandez A, Sumita D, Bobillo C, García MG, Gusmão L (2020) Paternal and maternal mutations in X-STRs: A GHEP-ISFG collaborative study, Forensic Sci Int Genet 46:1-7. https://doi.org/10.1016/j.fsigen.2020.102258

[7] Verogene ForenSeq™ DNA Signature Prep (2020) Reference Guide. D2018005. Rev. C. California 92121 U.S.A, San Diego.

[8] Parson W, Ballard D, Budowle B, Butler JM, Gettings KB, Gill P, Gusmao L, Hares DR, Irwin JA, King JL, Knijff PD, Morling N, Prinz M, Schneider PM, Neste CV, Willuweit S, Phillips C (2016) Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. Forensic Sci Int Genet 22:54–63 https://doi.org/10.1016/j.fsigen.2016.01.009

[9] Aguilar-Velázquez JA, Duran-Salazar MA, Cordoba-Mercado MF, Coronado-Avila CE, Salas-Salas O, Martínez-Cortes G, Casals F, Calafell F, Ramos-Gonzalez B, Rangel-Villalobos H (2022) Characterization of 58 STRs and 94 SNPs with the

ForenSeq™ DNA signature prep kit in Mexican-Mestizos from the Monterrey city (Northeast, Mexico). Mol Biol Rep 49(8):7601-7609. https://doi.org/10.1007/s11033-022-07575-y

[10] Cortés-Trujillo I, Zuñiga-Chiquette F, Ramos-González B, Chávez-Briones ML, Islas-González KL, Betancourt-Guerra DA, Peralta-Coria R, Martínez-Cortés G, Rangel-Villalobos H (2019) Allele and haplotype frequencies of 12 X-STRs in Mexican population. Forensic Sci Int Genet 38:11-13 https://doi.org/10.1016/j.fsigen.2018.10.012

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1099-1106
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Is it truly necessary to introduce the DYS570 and DYS576 into the PowerPlex® Fusion 6C panel to help in mixtures?

Francesco Ausania[1], Giulia Soldati[1], Chiara Saccardo[1], Dario Raniero[1], Domenico De Leo[1], Stefania Turrina[1]

[1] *Department of Diagnostics and Public Health, Unit of Forensic Medicine, University of Verona, Italy*

Corresponding Author: *Francesco Ausania · E-mail: francesco.ausania@univr.it*
Institute of Legal Medicine; Department of Diagnostic and Public Health · University of Verona – Italy
Policlinico G.B. Rossi, P.le L.A. Scuro · 37134 - Verona, Italy

Abstract
The ability to indicate whether a sample comes from a male or a female source and to distinguish the contributors of a mixture is particularly relevant for identifying the evidence of the victim and the perpetrator(s), especially in cases of sexual assault.
Since its introduction in 1994, the Amelogenin system has been the gold standard for gender determination also because it can be amplified with autosomal and gonosomal short tandem repeat markers.
Over the years, cases have been reported where deletion at position p11.2 of the Y-chromosome (Y-Chr) has resulted in failed amplification of AMEL Y, leading to gender misinterpretation. To prevent this occurrence, since 2014, the DYS391 locus, located at q11.21 on the long arm of the Y-Chr, has been introduced in autosomal STR kits as support for AMEL Y.
During a validation and population data acquisition study on 400 individuals in Northern Italy using the PowerPlex® Fusion 6C System, in a sample for which the AMEL Y deletion was already known, the non-amplification for the two Rapidly Mutating Y-STRs (RM Y-STRs) DYS570 and DYS576 was also revealed. The DNA from this sample was then subjected to further amplification with different X- and Y-chromosome amplification kits.
The findings raise the question of whether using Y-STRs in close proximity to the Amelogenin gene contributes to gender determination in mixtures or whether it would be more appropriate to use Y-STRs in positions less affected by deletion events.

Keywords
PowerPlex® Fusion 6C System; Y chromosome; Short arm deletions; Amelogenin

1. Introduction

Amelogenin represents a crucial marker for gender determination, particularly in the forensic context, as the accuracy of gender determination in biological samples is essential for investigative and probative purposes. Indeed, since 1994 [1], Amelogenin, which involves the amplification of AMEL X and AMEL Y, has been included in most forensic amplification kits, especially those that amplify for markers located on autosomal chromosomes. As there have been increasing reports in the Literature over the years of AMEL Y deletions at position p11.2 of the short arm of the Y chromosome [2-10], since 2014, a Y-chromosome (Y-chr) marker has been added to the amplification kits to support Amelogenin for gender determination: the DYS391, which is a stable locus with a relatively narrow allele range [11].

One of the most widely used autosomal short tandem repeats (aSTR) amplification kits in forensic genetics laboratories is the PowerPlex® Fusion 6C System (henceforth Fusion 6C), which allows the co-amplification of 27 loci, including the new 20 CODIS Core loci, Amelogenin, and three Y-STR (DYS391, DYS570, and DYS576). The first Y-STR was intended to confirm the gender of the sample revealed by Amelogenin, whereas the introduction of the two rapidly mutating Y-STRs (DYS576 and DYS570) is supportive in determining the number of contributors within a mixture [12]. Because these two loci are located close to the Amelogenin at position p11.2 (Figure 1), any deletion phenomena could compromise their original functionality within the Fusion 6C and thus not provide the necessary support for determining the number of contributors in a mixture.

This study raised a question concerning the effective contribution of the two RM Y-STRs loci in determining the gender and number of contributors in a mixture.

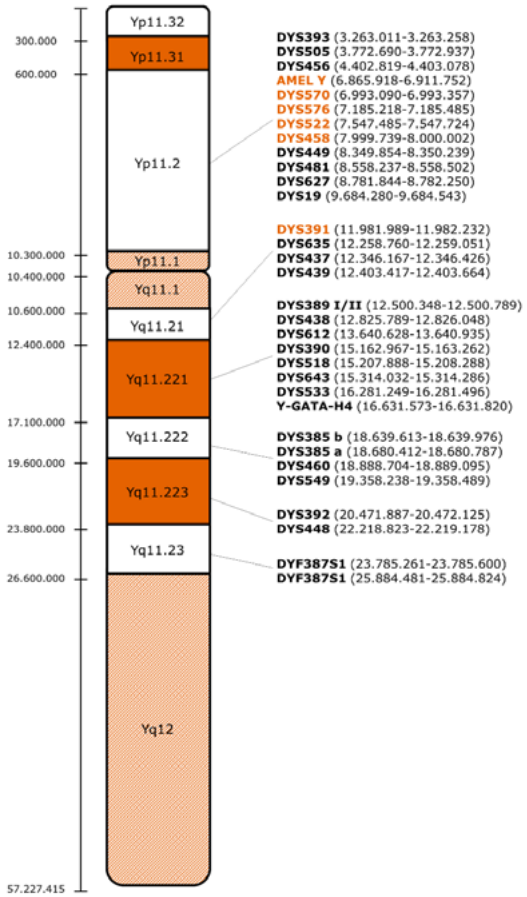


Figure 1. Y chromosome representation: all markers included in the amplification kits used for the analysis of this study were reported, together with their specific locations

2. Materials and methods

To perform an internal validation and population data acquisition study, the genomic DNA from 400 subjects (230 males and 170 females) from Northern Italy was isolated using the QIAamp DNA Mini Kit. The DNA was quantified with the Qubit Fluorimeter using the Qubit dsDNA HS Assay Kit and normalized to 1 ng/µl. The genomic DNA was then amplified with the PowerPlex® Fusion 6C System.

The sample pool chosen for the validation study also included a sample for which deletions at the AMEL Y and DYS458 loci were already known [5]; exclusively

for this sample, the genomic DNA was amplified with five additional kits, repeating each analysis twice: PowerPlex® Y23 System, Yfiler™ Plus PCR Amplification Kit, Investigator® Argus Y-28 QS Kit, Investigator Argus X-12 QS Kit, and ForenSeq DNA Signature Prep Kit, henceforth referred to as PowerPlex Y23, Yfiler Plus, Argus Y-28, Argus X-12, and FSSP, respectively.

Amplicon genotyping was performed using the SeqStudio™ Genetic Analyzer for HID equipped with GeneMapper® ID-X v1.6.

3. Results

In the validation and population data acquisition study, the typing of the 400 DNA samples amplified with the Fusion 6C produced complete and correctly typed profiles; the only exception was one sample, for which AMEL Y and DYS458 deletions were already known [5], which showed deletions at AMEL Y, DYS576, and DYS570 loci (Figure 2A). The three repetitions, carried out on the sample to assess the reproducibility of the data, always showed the same result.

DNA extracted from the same subject was amplified with other Y- and X-chromosome amplification kits to check for further deletions.

The typing conducted with the PowerPlex Y23, Yfiler Plus, and Argus Y-28, confirmed the deletions in the DYS576 and DYS570 markers, and the additional deletion at the DYS458 locus was again detected (Figure 2B).

The typing conducted with the Argus X-12 confirmed the AMEL Y deletion and showed no other deletions on the X chromosome (Figure 2C).

The typing conducted with the FSSP confirmed the deletions in the AMEL Y, DYS576, and DYS570 markers, and an additional deletion at the DYS522 locus was detected (Figure 2C).

The analysis with each kit was repeated twice, and the same genetic profile was always confirmed.

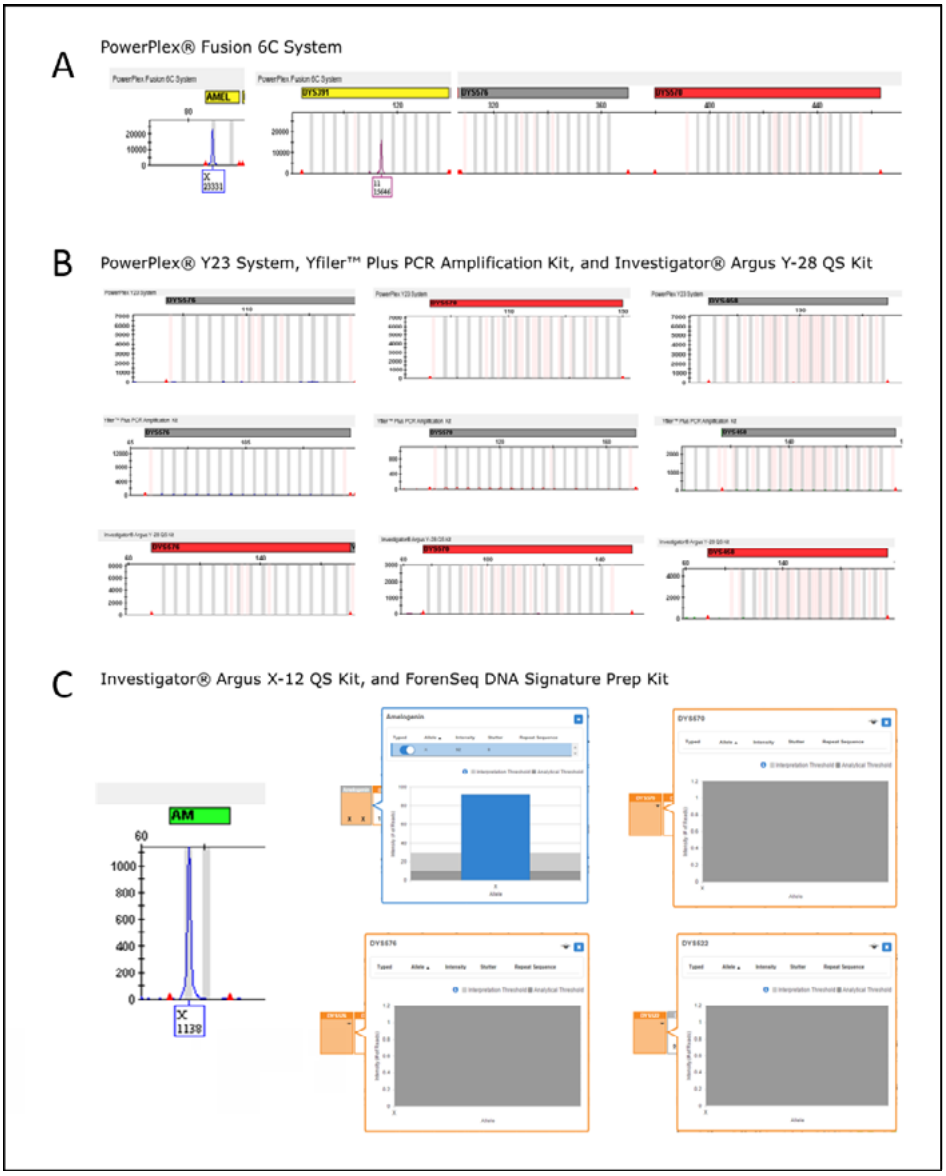


Figure 2. Electropherograms details of PowerPlex® Fusion 6C System, PowerPlex® Y23 System, Yfiler™ Plus PCR Amplification Kit, Investigator® Argus Y-28 QS Kit, Investigator Argus X-12 QS Kit, and ForenSeq DNA Signature Prep Kit. Deletions were found at AMEL Y, DYS570, DYS576, DYS458, and DYS522.

4. Discussion

From the preliminary analysis of the electropherogram obtained by DNA amplification with the Fusion 6C kit, the test sample appeared to belong to a female individual. This was inferred from the absence of alleles for AMEL Y and the two RM Y-STR markers, while the single peak detected in DYS391 could be considered an artefact, such as a spike. However, subsequent analysis with other Y-chr amplification kits confirmed that the sample belonged to a male individual. The latter had a significant deletion at position p11.2, evidenced by the absence of AMEL Y and four other nearby markers. Although deletions in the short arm of the Y-chr at position p11.2 are rare occurrences, they have been documented repeatedly in the literature and must be carefully considered. In light of the findings, the efficacy of the markers DYS576 and DYS570, included in the Fusion 6C kit, for determining sex and number of contributors in a mix can be questioned. Therefore, it would be appropriate to consider the inclusion of markers located further away from the Amelogenin locus, characterized by more excellent stability and not affected by deletion phenomena, to ensure greater consistency of results. This approach was adopted by introducing the DYS391 marker, located at q11.21, on the long arm of the Y-chr, which is significantly less prone to deletion phenomena.

5. Conclusion

This study used different amplification kits to analyze a sample with known deletions at the AMEL Y and DYS458 markers. The results revealed deletions also at the markers DYS570, DYS576, and DYS522. These markers are all located near the Amelogenin locus, and the latter's absence can influence the presence of deletions in adjacent markers, causing a multi-locus deletion. Although cases such as the one illustrated in this study are rare, they are documented in the Literature, and therefore, the efficacy of the DYS570 and DYS576 markers for determining the sex and number of contributors in a mixture may be questioned. Indeed, assigning the wrong sex to a sample may have severe implications in forensic or prenatal diagnosis cases. The Authors suggest that it would be more appropriate to include in the PowerPlex® Fusion 6C System, in support of the autosomal markers and DYS391, two Y-STR markers that are more stable and positioned away from the Amelogenin locus, preferably on the long arm of the Y chromosome, which is characterized by greater stability and less susceptibility to deletions. This approach could better meet forensic requirements, ensuring greater consistency of results.

6. Acknowledgements

None

7. Conflict of interest statement

None

8. References

[1] Mannucci, A., Sullivan, K. M., Ivanov, P. L., & Gill, P. (1994). Forensic application of a rapid and quantitative DNA sex test by amplification of the X-Y homologous gene amelogenin. International journal of legal medicine, 106(4), 190–193. https://doi.org/10.1007/BF01371335

[2] Lattanzi, W., Di Giacomo, M. C., Lenato, G. M., Chimienti, G., Voglino, G., Resta, N., Pepe, G., & Guanti, G. (2005). A large interstitial deletion encompassing the amelogenin gene on the short arm of the Y chromosome. Human genetics, 116(5), 395–401. https://doi.org/10.1007/s00439-004-1238-z

[3] Yong, R. Y., Gan, L. S., Chang, Y. M., & Yap, E. P. (2007). Molecular characterization of a polymorphic 3-Mb deletion at chromosome Yp11.2 containing the AMELY locus in Singapore and Malaysia populations. Human genetics, 122(3-4), 237–249. https://doi.org/10.1007/s00439-007-0389-0

[4] Kumagai, R., Sasaki, Y., Tokuta, T., Biwasaka, H., & Aoki, Y. (2008). DNA analysis of family members with deletion in Yp11.2 region containing amelogenin locus. Legal medicine (Tokyo, Japan), 10(1), 39–42. https://doi.org/10.1016/j.legalmed.2007.05.009

[5] Turrina S., Filippini G., De Leo D., (2009). Evaluation of deleted region from Yp11.2 of two amelogenin negative related males. Forensic Science International: Genetics Supplement Series, 2(1), 240-241. https://doi.org/10.1016/j.fsigss.2009.08.157

[6] Takayama, T., Takada, N., Suzuki, R., Nagaoka, S., Watanabe, Y., Kumagai, R., Aoki, Y., & Butler, J. M. (2009). Determination of deleted regions from Yp11.2 of an amelogenin negative male. Legal medicine (Tokyo, Japan), 11 Suppl 1, S578–S580. https://doi.org/10.1016/j.legalmed.2009.01.049

[7] Kumagai, R., Sasaki, Y., Tokuta, T., Biwasaka, H., Matsusue, A., Aoki, Y., & Dewa, K. (2010). Distinct breakpoints in two cases with deletion in the Yp11.2 region in Japanese population. Human genetics, 127(5), 537–543. https://doi.org/10.1007/s00439-010-0794-7

[8] Chen, W., Wu, W., Cheng, J., Zhang, Y., Chen, Y., & Sun, H. (2014). Detection of the deletion on Yp11.2 in a Chinese population. Forensic science international. Genetics, 8(1), 73–79. https://doi.org/10.1016/j.fsigen.2013.07.003

[9] Cheng, J. B., Liu, Q., Long, F., Huang, D. X., & Yi, S. H. (2019). Analysis of the Yp11.2 Deletion Region of Phenotypically Normal Males with an AMELY-Null Allele in the Chinese Han Population. Genetic testing and molecular biomarkers, 23(5), 359–362. https://doi.org/10.1089/gtmb.2018.0231

[10] Pang, Q., Lin, Q., Wang, D., Sun, Z., & Wang, J. (2021). Molecular characterization of the Yp11.2 region deletion in the Chinese Han population. International journal of legal medicine, 135(4), 1351–1358. https://doi.org/10.1007/s00414-021-02596-x

[11] Butler J.M. (2013) Introductory Remarks. New Autosomal and Y-STR Loci and Kits: Making Data Driven Decisions. ISHI Workshop on New Loci and Kits (Atlanta, GA)

[12] PowerPlex® Fusion 6C System Brochure - Performance Packed (Promega). Website: https://ita.promega.com/resources/pubhub/ebrochures/powerplex-fusion-6c-system/

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1107-1114
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Enhancing Signal Precision in SNaPshot™ Multiplex Kit Genotyping of IrisPlex Targets

Saurabh Patil[1], Bhargav Patel[1,2,*]

[1]*Laboratory of Forensic Biology and Biotechnology, School of Forensic Science, National Forensic Sciences University (NFSU), Gandhinagar, Gujarat, India*

[2]*Center of Excellence in DNA Forensics (CoE-DNA), National Forensic Sciences University (NFSU), Gandhinagar, Gujarat, India*

**Corresponding author: Dr. Bhargav C. Patel, School of Forensic Science, National Forensic Sciences University (NFSU), Gandhinagar, Gujarat, India. Email: bhargav.patel@nfsu.ac.in*

Abstract

Accomplishing specificity in multiplex polymerase chain reactions (PCR) and single base extension (SBE) reactions poses a significant challenge. Despite designing primers meticulously, residual non-specific nucleotide interactions often persist, impacting the accuracy of genotyping results. This study addresses the issue of non-specific amplifications in multiplex PCR and SBE reactions used for genotyping single nucleotide polymorphisms (SNPs) in the IrisPlex tool. Established protocol for genotyping IrisPlex targets using the SNaPshot™ multiplex assay resulted in non-specific electropherogram peaks. Potential signal peaks and non-specific peaks were identified by comparing electropherograms of singleplex and multiplex reactions. Non-specific peaks were found to be primer dimers formed by interactions between primers. For optimization of genotyping, touchdown multiplex PCR (67 °C to 60 °C), varying PCR primer concentrations (0.05 µM to 0.4 µM), and ammonium sulfate (AS) (0 mM, 50 mM, 80 mM, 100 mM) to the SBE reaction were carried out. Non-specific peaks were reduced from 19 to 5, while signal peaks of all 6 SNPs were identified. Thus, a significant reduction in non-specific peaks was achieved, while maintaining sufficient signal peak heights leading to enhanced specificity in the genotyping workflow. These findings suggest a successful optimization of the genotyping workflow, thereby enhancing the reliability of IrisPlex genotyping.

Keywords

IrisPlex, Single base extension, SNaPshot, Multiplex, Primer dimer, Touchdown PCR

1. Introduction

Multiplex PCR involves the amplification of more than one target sequence by respective oligonucleotide pairs. However, interactions of these oligonucleotides sometimes result in non-specific amplifications, thereby hampering the specificity of the reaction. Non-specificity primarily occurs due to unintended primer binding to non-target DNA sequences or interactions between primers, resulting in non-specific amplification. Careful design of multiplex primers can limit the non-specific amplification to a great extent, however, some may persist [1]. SBE reaction involves SNP genotyping by the addition of dideoxynucleotide triphosphate (ddNTP) to the 3' end of the oligonucleotide which is complementary to the target sequence. Similar to PCR, multiplex SBE reactions can also be affected by non-specificity [2]. In such cases, the optimization of PCR &/or SBE conditions and reagents is warranted.

IrisPlex is a web-based tool that predicts eye color based on genotypes of 6 SNPs. Genotyping of these SNPs is mainly carried out by SNaPshot™ chemistry [3, 4, 5]. However, non-specificity is one of the common issues associated with this chemistry [2]. It causes ambiguity in genotyping interpretation. This study aims to optimize SNaPshot™ kit-based IrisPlex SNPs' genotyping to resolve non-specificity and ambiguity in genotyping interpretation.

2. Material and methods

2.1. Sample collection

In this study, peripheral blood samples were collected from healthy individuals with informed consent following the declaration of Helsinki. The collected blood samples were stored in EDTA vacutainers at 4 °C till further processing. This study was approved by the Institute Ethics Committee (IEC) of National Forensic Sciences University, Gandhinagar, India vide letter no. NFSU/SDSR/IEC/Certificate/273/21.

2.2. DNA extraction and quantitation

DNA was extracted using the DNeasy® Blood and Tissue kit (Qiagen) as per the manufacturer's protocol. Extracted DNA was quantified by UV-visible spectroscopy using Agilent BioTek Cytation 5 Cell Imaging Multimode Reader.

2.3. Multiplex PCR, SBE, and amplified product cleanup

For genotyping IrisPlex's SNPs, PCR and SBE reactions followed the previously described protocol and primers [3]. Multiplex PCR was optimized by modulating cycling conditions (Touchdown PCR: 95 °C for 5 mins → 33 cycles of 95 °C for 30 s, 67 °C to 60 °C (reduction of 1 °C per cycle till 60 °C) for 15 s, 72 °C for 15 s → 72 °C for 5 mins → 4 °C). PCR reaction volume was optimized to 20 µl. In it, 0.5 unit (U) of AmpliTaq Gold DNA Polymerase (ThermoFisher Scientific), 1X polymerase buffer, $MgCl_2$ (0.25 mM), dNTPs (0.2 mM), PCR primer concentrations (0.05 µM to 0.4 µM), DNA (1 ng), and nuclease-free water (NFW; for volume makeup). PCR products (3 µl) were cleaned by Exonuclease I (ExoI; 10 U) and Shrimp Alkaline Phosphatase (SAP; 1 U) (37 °C for 1 hour → 85 °C for 15 mins). For multiplex SBE, the reaction volume was optimized to 5 µl, in which cleaned PCR product (1 µl), SNaPshot™ reaction mix (1 µl), SBE primer concentrations (0.1 µM to 0.3 µM), and AS (0 mM, 50 mM, 80 mM, 100 mM) were added. AS was added to the SBE primer mix. SBE cycling conditions were set as described earlier [3]. SBE products were cleaned by SAP (37 °C for 1 hour → 85 °C for 15 mins). PCR, SBE reaction, and Exo-SAP cleanup were carried out on the ProFlex™ PCR system (ThermoFisher Scientific).

2.4. Capillary electrophoresis (CE) and analysis

Cleaned SBE products were separated and detected by CE on the ABI 3500 Genetic Analyzer (Life Technologies) using 1 µl cleaned SBE product, 8.7 µl Hi-Di™ formamide, and 0.3 µl GeneScan™ 120 LIZ® Size Standard (ThermoFisher Scientific). For CE, optimized instrument module parameters were: Pre-run (10 kV/ 200 s), Injection (1.6 kV/ 8 s), Run (10 kV/ 900 s), and Data Delay (200 ms). CE results (.fsa files) were analyzed on GeneMapper™ ID-X v1.6 software. Plots were created using the ggplot2 package of the R programming language for quantitative comparison of electropherogram peaks [6].

3. Results

Application of IrisPlex SNP genotyping protocol [3], resulted in 25 peaks in the electropherogram, much more than the expected maximum of 12 peaks (Fig. 1A). This suggested the presence of non-specific interactions in the reaction. Comparison with no-DNA template and singleplex reaction electropherograms identified non-specific peaks and also signal peaks corresponding to individual SNPs.

Interestingly all non-specific peaks were also present in the no-DNA template reaction. In this reaction, as only amplification of primer interactions can occur, the non-specific peaks observed seemed to be primer dimers. Peak numbers 2, 10, 15, and 23 & 24 corresponded to rs1800407, rs16891982, rs12913832, and rs12896399, respectively. To address non-specificity, optimization of the genotyping protocol was carried out. Touchdown PCR with a primer concentration of 0.1 µM each reduced non-specific peaks from 19 to 5 (Fig. 1A).

However, categorizing peaks 11, 12, and 20 as either signal peaks or primer dimers remained challenging as these peaks were present in both DNA template and no-DNA template reactions. AS was added in the SBE step to resolve the ambiguity in the peak interpretation. A progressive reduction of peak heights with an increase in AS concentration was observed (Fig. 1B). Average signal peak heights varied from 1986.6 relative fluorescence unit (RFU) in 0 mM AS to 329 RFU in 100 mM AS (Fig. 2A). Peak number 12 diminished from 50 mM concentration onwards, while the intensity of peaks 11 and 20 were still above the analytical threshold (50 RFU) even at 100 mM concentration of AS. This is important as only the peaks above the analytical threshold can be considered for genotyping. These results suggest peaks 11 and 20 correspond to rs12203592 and rs1393350, respectively, while peak 12 was a non-specific peak. In the 100 mM AS reaction, the intensity of the signal peak of rs1800407 fell below the analytical threshold, making it unsuitable for genotyping (Fig. 2B).

4. Discussion

IrisPlex system predicts iris color (brown, blue, or intermediate) using genotypes from 6 SNPs, with genotyping typically performed via SNaPshot™ Multiplex chemistry [3]. In this method, fluorescently labeled ddNTPs extend oligonucleotide probes complementary to the target region, revealing the SNP's genotype. This is a widely used chemistry, however, non-specificity is a common problem associated with it [2]. Avoiding non-specificity in this assay is important as it may mask signal peaks. It may also prevent the detection of novel or rare alleles of SNPs.

In this study, we report the non-specificity associated with the SNaPshot™ based genotyping of IrisPlex SNPs and the optimization of this protocol to minimize the associated non-specificity. Primer interactions and non-specific binding of primer to target DNA cause non-specificity in PCR and SBE [1]. Touchdown PCR is highly effective in limiting such non-specific amplifications. In touchdown

PCR, the starting annealing temperature is set higher than the actual annealing temperature. It is then gradually reduced in each cycle till a lower permissive temperature (approximately T_m - 5 °C). Higher temperatures disrupt weaker interactions between primers and between primers and the template, thereby promoting the amplification of more stable interactions, primarily those between primers and their specific target sequences [9]. This is evident in the current study where touchdown PCR significantly reduced non-specific peaks while largely retaining the signal peaks (Fig. 2B). Touchdown from 67 °C to 60 °C was found optimal in the present study. As the addition of AS to the SBE primer mixture significantly reduced primer dimers in a previous study [8], AS was added to the SBE reaction. In conjunction with touch-down PCR, AS reduced non-specificity when added to SBE. It also helped determine the signal peaks of rs12203592 and rs1393350. NH_4^+ ions of AS destabilize hydrogen bonding between nucleotides. As a result, weaker interactions break and non-specific amplifications are avoided [9]. However, this action also affects other more stable interactions, which explains the reduction in signal peak intensities (Fig. 3B). Overall, 80 mM AS in SBE seemed optimum as non-specific peaks were minimal with all signal peaks retained, thereby maximizing genotyping efficiency as compared to other concentrations.

To summarize, touchdown PCR coupled with AS in SBE reduced non-specificity while maintaining the robustness of the genotyping of IrisPlex SNPs. In this study, the genotypes of the samples used were unknown. This optimized protocol needs to be further validated with DNA samples or standards with known genotypes of IrisPlex SNPs to confirm the accuracy of genotyping.

5. Conclusion

Our results demonstrate effective measures to minimize non-specificity in SNaPshot™ Multiplex Kit-based genotyping of IrisPlex SNPs. In this regard, touchdown PCR and AS as an additive in SBE emerged as a useful combination. This approach improves the reliability of IrisPlex SNP genotyping by significantly reducing non-specific peaks, making it a more robust tool for predicting iris color in forensic and research settings. The combination of touchdown PCR and AS in SBE can potentially be applied to other genotyping assays facing similar issues. Future work will focus on validating this protocol using DNA samples with known IrisPlex SNP genotypes to ensure the accuracy of the optimized genotyping method.

6. Acknowledgments

The authors are thankful to Harshit Ganatra, Molina Madhulika Ekka, Mohammad Alqaisi, Mohamed Eeman, and Subhojit Mitra for their valuable contributions to technical discussions, manuscript review, and sample collection. The authors are grateful to NFSU Gandhinagar for providing intramural research funding, and to the School of Forensic Science and CoE-DNA at NFSU Gandhinagar for providing laboratory facilities. SP thankfully acknowledges CSIR, Govt. of India for providing a Junior Research Fellowship (09/1149(16456)/2023-EMR-I).

7. Conflict of interest statement

None

8. References

[1] Markoulatos, P et al. "Multiplex polymerase chain reaction: a practical approach." Journal of clinical laboratory analysis vol. 16,1 (2002): 47-51. doi.org/10.1002/jcla.2058

[2] Fondevila, M et al. "Forensic SNP genotyping with SNaPshot: Technical considerations for the development and optimization of multiplexed SNP assays." Forensic science review vol. 29,1 (2017): 57-76.

[3] Walsh, Susan et al. "IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information." Forensic science international. Genetics vol. 5,3 (2011): 170-80. doi.org/10.1016/j.fsigen.2010.02.004

[4] Dario, Paulo et al. "Assessment of IrisPlex-based multiplex for eye and skin color prediction with application to a Portuguese population." International journal of legal medicine vol. 129,6 (2015): 1191-200. doi.org/10.1007/s00414-015-1248-5

[5] Kastelic, Vanja et al. "Prediction of eye color in the Slovenian population using the IrisPlex SNPs." Croatian medical journal vol. 54,4 (2013): 381-6. doi.org/10.3325/cmj.2013.54.381

[6] Wickham H (2016). "ggplot2: Elegant Graphics for Data Analysis." Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

[7] Korbie, Darren J, and John S Mattick. "Touchdown PCR for increased specificity and sensitivity in PCR amplification." Nature protocols vol. 3,9 (2008): 1452-6. doi.org/10.1038/nprot.2008.133

[8] Sanchez, Juan J et al. "A multiplex assay with 52 single nucleotide polymorphisms for human identification." Electrophoresis vol. 27,9 (2006): 1713-24. doi.org/10.1002/elps.200500671

[9] Van Pelt-Verkuil, Elizabeth, et al. "Principles and Technical Aspects of PCR Amplification." Springer eBooks, 2008, https://doi.org/10.1007/978-1-4020-6241-4.

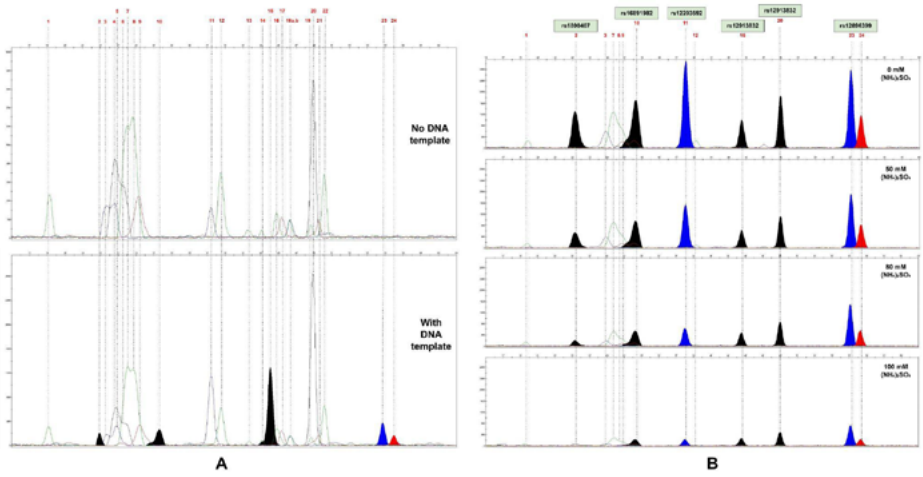
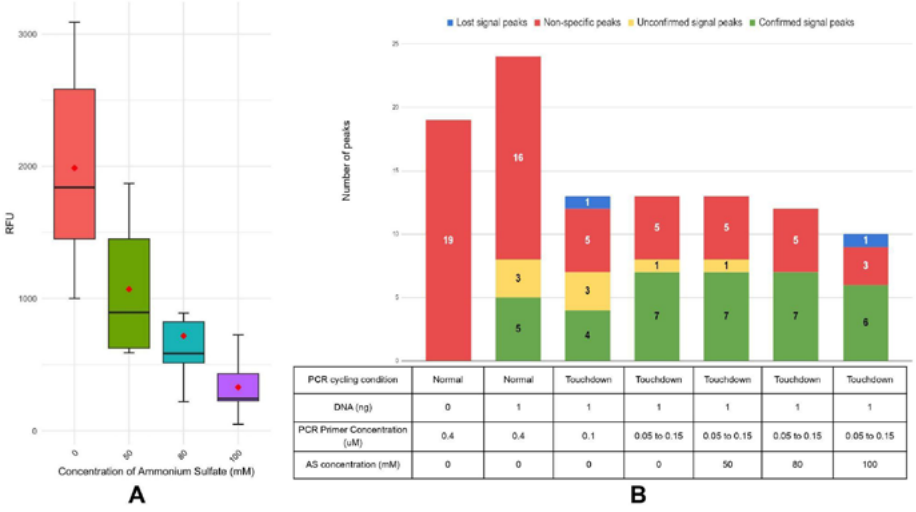


Figure 1. (A) Electropherogram comparison of SNaPshot™-based IrisPlex genotyping reactions without a DNA template (top) and with a DNA template (bottom). (B) Electropherograms of IrisPlex genotyping reactions with varying AS [$(NH_4)_2SO_4$] concentrations. Dotted lines map corresponding peaks between electropherograms, where filled peaks represent signal peaks and the others indicate non-specific peaks.



PCR cycling condition	Normal	Normal	Touchdown	Touchdown	Touchdown	Touchdown	Touchdown
DNA (ng)	0	1	1	1	1	1	1
PCR Primer Concentration (uM)	0.4	0.4	0.1	0.05 to 0.15	0.05 to 0.15	0.05 to 0.15	0.05 to 0.15
AS concentration (mM)	0	0	0	0	50	80	100

Figure 2. Quantitative analysis of electropherograms of IrisPlex genotyping reactions. (A) Average signal peak heights (red dots) across IrisPlex SNPs under different concentrations of AS. (B) Comparison of peaks (non-specific peaks, confirmed signal peaks, lost signal peaks, unconfirmed signal peaks) across different optimization conditions (PCR cycling condition, DNA amount, PCR primer concentration, AS concentration).

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1115-1120
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Determination of Human Facial Morphology: Eye and Forehead Structure

Sumeyye Zulal Simsek[1*], Ozlem Bulbul[1], Gonul Filoglu[1]

[1]*Istanbul University-Cerrahpasa, Institute of Forensic Sciences and Legal Medicine, Department of Science, Buyukcekmece, Istanbul, Turkey.*

** Corresponding author at: Istanbul University-Cerrahpasa, Institute of Forensic Sciences and Legal Medicine, Department of Science, Buyukcekmece, Istanbul, Turkey, sumeyye.zulal@gmail.com*

Abstract
Forensic DNA phenotyping (FDP) holds significant potential in criminal investigations where traditional methods, such as eyewitness accounts or conventional DNA profiling, may be insufficient. Genetic polymorphism defines phenotypic characteristics, with genome-wide association studies (GWAS) identifying loci that affect facial morphology. Research on facial features, including those related to craniofacial anomalies and hereditary diseases, highlights the importance of the eyes, forehead, nose, lips, cheeks, and cheekbones in individual identification. Single nucleotide polymorphisms (SNPs) affecting these regions help define facial structure and anomalies. These anatomical regions vary among individuals within and across populations. It has been determined that some of the identified SNP markers are effective in shaping facial structure, while others influence the formation of facial anomalies. The height, depth, and length of the eye are important in determining eye phenotype. The main objective of this study is to identify SNP loci that play a crucial role in determining the shape of the eyes and forehead, highlighting individual variations, which can aid in identifying suspects.

Keywords
Facial morphology, FDP, forensic genetics, SNP.

1. Introduction

Forensic DNA phenotyping (FDP) is an advanced technique used in forensic science to predict an individual's physical traits based solely on DNA evidence. Unlike

traditional forensic DNA analysis, which relies on comparing DNA profiles to known individuals or databases, FDP focuses on identifying genetic markers associated with phenotypic traits such as hair colour, eye colour, skin pigmentation, and even facial structure. This method leverages single nucleotide polymorphisms (SNPs) that are linked to observable characteristics, providing valuable information when no reference samples or suspects are available. FDP has become particularly useful in cases where DNA is too degraded for conventional methods or when investigative leads are lacking. By offering clues about a suspect's appearance or biogeographical ancestry, FDP assists law enforcement in narrowing down potential suspects and generating investigative leads, though it remains a probabilistic tool rather than a definitive form of identification [1], [2].

Genome-wide association studies (GWAS) have identified the loci involved in shaping facial morphology [1]. These studies have helped uncover genetic markers by focusing on various facial features, particularly those associated with craniofacial anomalies such as cleft lip and palate syndrome, as well as hereditary conditions like Down Syndrome [3]the genetic variation that underpins normal human facial appearance is still largely elusive. Recent development of novel digital methods for capturing the complexity of craniofacial morphology in conjunction with high-throughput genotyping methods, show great promise for unravelling the genetic basis of such a complex trait.\n\nAs a part of our efforts on detecting genomic variants affecting normal craniofacial appearance, we have implemented a candidate gene approach by selecting 1,201 single nucleotide polymorphisms (SNPs. Key facial regions that play a role in defining an individual's appearance include the eyes, forehead width, nose, lips, cheeks, and cheekbones. These features vary not only among individuals but also across different populations. While facial morphology shares a common aetiological basis, genetic polymorphisms result in significant differences among populations, such as Caucasian, Asian, and African groups. These genetic differences, or SNP markers, are scattered throughout the genome [4]. Some SNPs influence the normal development of facial structure, while others contribute to the formation of facial anomalies [5]–[7]allowing for an open-ended description of facial variation while preserving statistical power. In a sample of 2,329 persons of European ancestry, we identified 38 loci, 15 of which replicated in an independent European sample (n = 1,719. Research has shown that the development of these features is linked to SNP points on genes such as *OSR1-WDR35*, *HOXD1-MTX2*, *WDR27*, and *SOX9* [5], [8]–[10]only a small number of genes related to face morphology have been identified to date. Here, we performed a two-stage genome-wide association study (GWAS.

2. Material studied, methods, techniques

Blood and/or swab samples were collected from 100 unrelated individuals aged 20–45 who consented to DNA analysis. None of the individuals had experienced facial trauma or undergone aesthetic procedures. Standardised photographic documentation was taken under optimal conditions, including frontal and lateral views. Facial measurements were conducted using ImageJ2 software (Figure 1). A panel of 19 highly correlated SNP loci associated with eye and forehead morphology was selected following a literature review. The SNaPshot® methodology was used to analyse the panel.

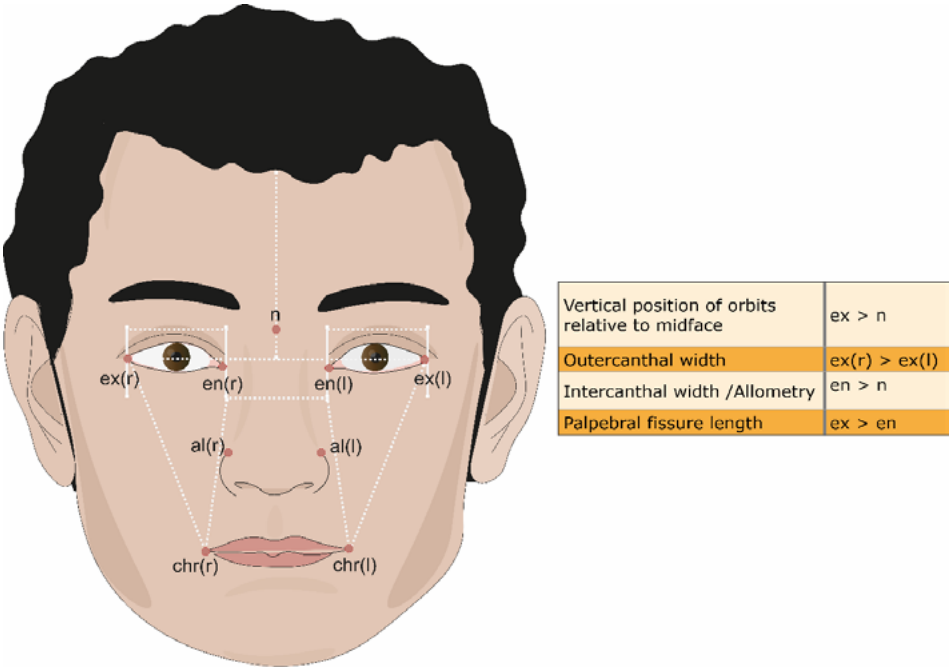


Figure 1. Measurement of eye and forehead structure

3. Results and Discussion

As a preliminary result of the study, the optimisation of two panels related to the eye and forehead regions was successfully completed (Figure 2). In the validation

study, the minimum DNA concentration was determined to be 1 ng/µl. The optimisation and validation of the 19-SNP panel were successfully carried out. Genotype analyses were conducted on samples from 100 volunteers.

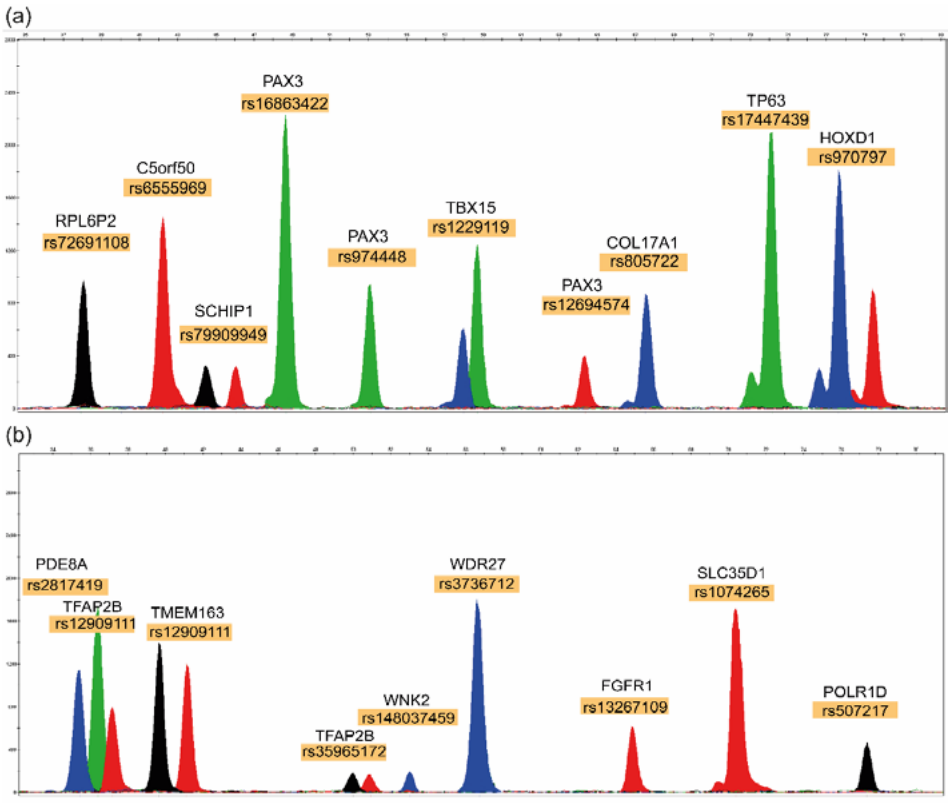


Figure 2. Electropherogram of the 19-SNP Panel showing Multiplex A (10 SNPs) and Multiplex B (9 SNPs)

4. Conclusion

Moving forward, we will apply multinomial logistic regression (MLR) to predict morphological traits from population samples, enhancing our model for generating molecular facial composites. The population study, which will be conducted using the developed panel, aims to create a model for generating a molecular facial composite in cases where suspects are absent and missing persons need to

be identified. This innovative approach holds potential for aiding investigations in such cases and in the detection of missing persons, offering a promising avenue for enhancing forensic identification methods.

5. Acknowledgments

Authors would like to thank all the volunteers for their participation in this study. The presentation of this work at the 30th Congress of the ISFG was fully supported by GENETIKS Global Genetic Diseases Evaluation Center and Laboratories.

6. Conflict of interest statement

None.

7. References

[1] L. A. Marano and C. Fridman, "DNA phenotyping: current application in forensic science," *Res. Reports Forensic Med. Sci.*, vol. Volume 9, pp. 1–8, Feb. 2019, doi.org/10.2147/RRFMS.S164090.

[2] M. Kayser, W. Branicki, W. Parson, and C. Phillips, "Recent advances in Forensic DNA Phenotyping of appearance, ancestry and age," *Forensic Sci. Int. Genet.*, vol. 65, p. 102870, Jul. 2023, doi.org/10.1016/j.fsigen.2023.102870.

[3] M. Barash, P. Bayer, and A. van Daal, "Candidate gene scan for Single Nucleotide Polymorphisms involved in the determination of normal variability in human craniofacial morphology," *bioRxiv*, p. 060814, 2016, doi.org/10.1101/060814.

[4] S. Richmond, L. J. Howe, S. Lewis, E. Stergiakouli, and A. Zhurov, "Facial genetics: A brief overview," *Front. Genet.*, vol. 9, no. OCT, Oct. 2018, doi.org/10.3389/fgene.2018.00462.

[5] P. Claes *et al.*, "Genome-wide mapping of global-to-local genetic effects on human facial shape," *Nat. Genet.*, vol. 50, no. 3, pp. 414–423, Mar. 2018, doi.org/10.1038/s41588-018-0057-4.

[6] P. Claes *et al.*, "Modeling 3D Facial Shape from DNA," *PLoS Genet.*, vol. 10, no. 3, p. e1004224, Mar. 2014, doi.org/10.1371/journal.pgen.1004224.

[7] B. Mark, B. Philipp E, and D. Angela van, "Identification of the Single Nucleotide Polymorphisms Affecting Normal Phenotypic Variability in Human Craniofacial Morphology Using Candidate Gene Approach," *J. Genet. Genome Res.*, vol. 5, no. 1, 2018, doi.org/10.23937/2378-3648/1410041.

[8] S. Cha *et al.*, "Identification of five novel genetic loci related to facial morphology by genome-wide association studies," *BMC Genomics*, vol. 19, no. 1, pp. 1–17, 2018, doi.org/10.1186/s12864-018-4865-9.

[9] J. R. Shaffer *et al.*, "Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology," *PLoS Genet.*, vol. 12, no. 8, pp. 1–21, 2016, doi.org/10.1371/journal.pgen.1006149.

[10] F. Liu *et al.*, "A Genome-Wide Association Study Identifies Five Loci Influencing Facial Morphology in Europeans," *PLoS Genet.*, vol. 8, no. 9, 2012, doi.org/10.1371/journal.pgen.1002932.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1121-1128
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Inferencing Fingerprints TsD Based on Copy Number Differences of Multiple mtDNA markers

Chong Chen[1], Ying Zhu[1], Minzhu Zhao[1], Qi Wang[1], Kai Yu[1], Yunhui Liu[1], Yufan Meng[1], Xiaoqiu Chen[1], Xiaoya Lu[1], Jianbo Li[1*]

[1]*Department of Forensic Medicine, Faculty of Basic Medical Science, Chongqing Medical University, Chongqing, China.*

**Corresponding author at: Department of Forensic Medicine, Faculty of Basic Medical Science, Chongqing Medical University, Chongqing, China. 00390@cqmu.edu.cn*

Abstract
Fingerprints are commonly encountered contact materials at crime scenes, and the determination of their time since deposition (TsD) is pivotal in establishing the timeline of the crime and finding suspects, which is a challenging issue urgently awaiting resolution in forensic science. However, traditional methods of fingerprint TsD inference have various limitations. Mitochondrial DNA (mtDNA), with its structurally stable and multicopy nature within cells, offers application advantages in trace and degraded biological samples. Longer gene fragments are more prone to degradation compared to shorter ones. Thus, by comparing the copy numbers of mtDNA markers with different lengths, we developed a quantitative index for fingerprint degradation. The findings indicate that mtDNA molecular markers have the potential to serve as powerful biomarkers for fingerprint TsD inference.

Keywords
Fingerprints; Time since deposition; mtDNA; Copy number variation; qPCR.

1. Introduction

Fingerprints are commonly found at crime scenes. Determining the time since deposition (TsD) of fingerprints assists in elucidating the timeline of a crime and suspect identification, etc. For example, suspects may concede to leaving their fingerprints at the scene but often claim the impressions were made at a time unrelated

to the criminal event, invoking this distinction for legal defense. Fingerprints have distinctive morphologies and are primarily composed of glandular secretions and epidermal cells. Previous studies for estimating the fingerprints TsD mainly focused on morphological characteristics, optical features and non-DNA components of fingerprints, facing limitations in accuracy, sensitivity, and specificity [1-5].

The contemporary forensic landscape demands more precise inference methods for fingerprint TsD inference. Mitochondrial genome has unique application advantages in trace and degraded samples due to its structural stability and high copy number in cells. The previous study has proved that mtDNA can be candidate molecular markers for blood TsD inference based on based on fluorescence method [6]. Longer DNA fragments tend to be more susceptible to degradation than shorter DNA fragments. Thus, in this study, we developed a set of mtDNA molecular markers with varying lengths detected by TaqMan probe method, evaluating their ratios against reference genes as degradation indexes, to accurately infer fingerprint TsD.

2. Material studied, methods, techniques

2.1 Preliminary screening mtDNA molecular markers and designing primers

Based on the CO1 gene in the mitochondrial genome sequence (NC_012920), mtDNA markers with different lengths were screened as TsD molecular markers. Primers and corresponding TaqMan MGB probe were designed using Oligo software v7.6. The species specificity of the candidate mtDNA primers were tested on NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi); nuclear mitochondrial DNA segments (NUMTs) were detected on USCS (https://genome.ucsc.edu/). The conservation of the mtDNA primers is verified in the mitochondrial genome sequences of Chinese populations. The primer amplification specificity and amplification efficiency of candidate mtDNA markers were further verify and optimize based on QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific, USA).

2.2 Sample preparation

In accordance with the medical ethics, the fingerprints of five unrelated individuals aged 24 to 25 years old from Chinese Han population were collected after signing informed consent. The personal information and health status of the participants were recorded. The volunteers showed that they pressed the clean

slide with three fingertips for 30 seconds and repeated twice. The slide was then placed in a cool, ventilated and sunlight-free environment (average temperature: 25°C; average humidity: 65%), and were collected at different time points (0 days, 3 weeks, 6 weeks...). With the volunteers' written consent, 2 mL of their peripheral blood was extracted for the preparation of standard samples.

2.3 Preparation of quantitative standards

Total DNA from peripheral blood was extracted using the MagaZorb® DNA Mini-Prep Kit (Promega, USA). The copy number of mtDNA were absolutely quantified by probe method on digital PCR platform (Bio-rad, USA). The components of the detection system mainly include ddPCR™ Supermix for Probes, mtDNA-specific primers, FAM/MGB probes, DNA templates and adding ultrapure water to the 20 μL system. Then, the above system was formed into droplets in the QX200 droplet generator, and transferred to 96-well plate for PCR amplification. Finally, QX200 Droplet Reader is used for droplet reading and data analysis.

2.4 DNA extraction from fingerprint samples and quantification of mtDNA markers

Use the double swab method to wipe the fingerprint sample on the glass slide, and use the magnetic bead method of the MagaZorb® DNA Mini-Prep Kit (Promega, USA) to extract the total DNA of the fingerprint. The quantification of the COI multiple fragments was further performed using TaqMan MGB probe method on the QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific, USA). The qPCR detection system of AceQ qPCR Probe Master Mix kit (Vazyme, China) is 10μL, including 5μL of Master Mix, 0.2μL of 10um forward/reverse primers, 0.1μL of 10um Taqman MGB probe, 1μL of DNA template, and 3.5μL of ddH2O. Each sample was tested in triplicate for each mtDNA marker. The cycle parameters of qPCR are: pre-denaturation at 95 °C for 5 minutes, followed by cycle reaction at 95 °C for 10 seconds and 60 °C for 30 seconds, 40 cycles.

2.5 Data analyses

Three molecular markers with different lengths (~60bp, ~140bp and ~350bp) distributed in the mitochondrial COI gene were selected, which were further marked as COI_{60bp}, COI_{140bp} and COI_{350bp}, respectively. The copy number of mtDNA was calculated according to the qPCR standard curve formula. The natural logarithms of the copy number ratios of different mtDNA genetic markers were used as the degradation index to evaluate the degradation degree of COI gene. The

calculation formulas were shown below.

$$Index^{a} = \ln (COI_{140bp}/COI_{60bp})$$

$$Index^{b} = \ln (COI_{350bp}/COI_{60bp})$$

3. Results

The COI_{60bp} molecular marker was used as the internal reference gene, and the COI_{140bp} and COI_{350bp} molecular markers exhibited no NUMTs. In Figure1, the melting curves of three candidate TsD markers were in the form of a single peak. The amplification efficiencies of the three TsD markers were all over 90%.

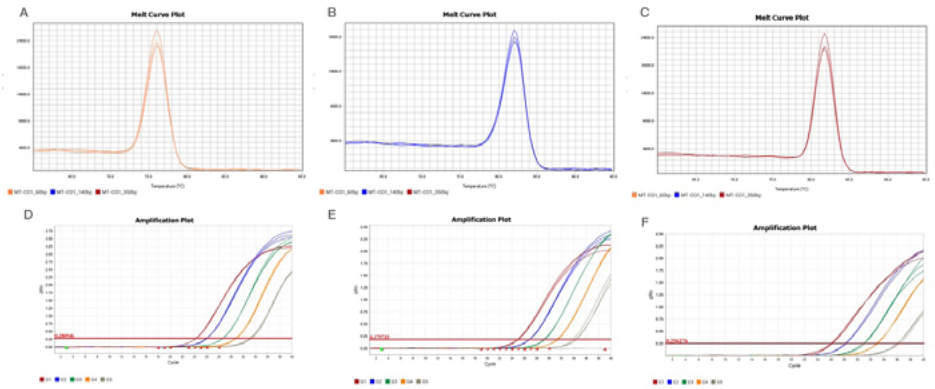


Figure 1. Evaluating amplification specificities of TsD primers based on Melting curves, (A) for COI_{60bp} primers, (B) for COI_{140bp} primers, (C) for COI_{350bp} primers. Detection of amplification efficiency of TsD primers based on 5-fold gradient dilution standards, (D) for COI_{60bp} primers, (E) for COI_{140bp} primers, (F) for COI_{350bp} primers.

The total DNA of peripheral blood was absolutely quantified on the ddPCR platform, which was further utilized as standards for qPCR quantification. As shown in Figure 2A & 2B, the concentration of the standard was 1160 copy/μl. Subsequently, after extracting DNA from the fingerprint samples at different time points, we employed qPCR probe method to quantify three TsD mtDNA markers. In Figure 2C, the concentration of mitochondrial genes in aged fingerprints (21 days) were found to be lower compared to fresh fingerprints. In Figures 2D & 2E, the degradation indexes of some fingerprint samples showed significant decreasing trends. Analyses

based on the two degradation indexes indicated a discernible degradation pattern in some fingerprints after 21 days, with the degradation rates for the Index[b] notably surpassing that of Index[a].

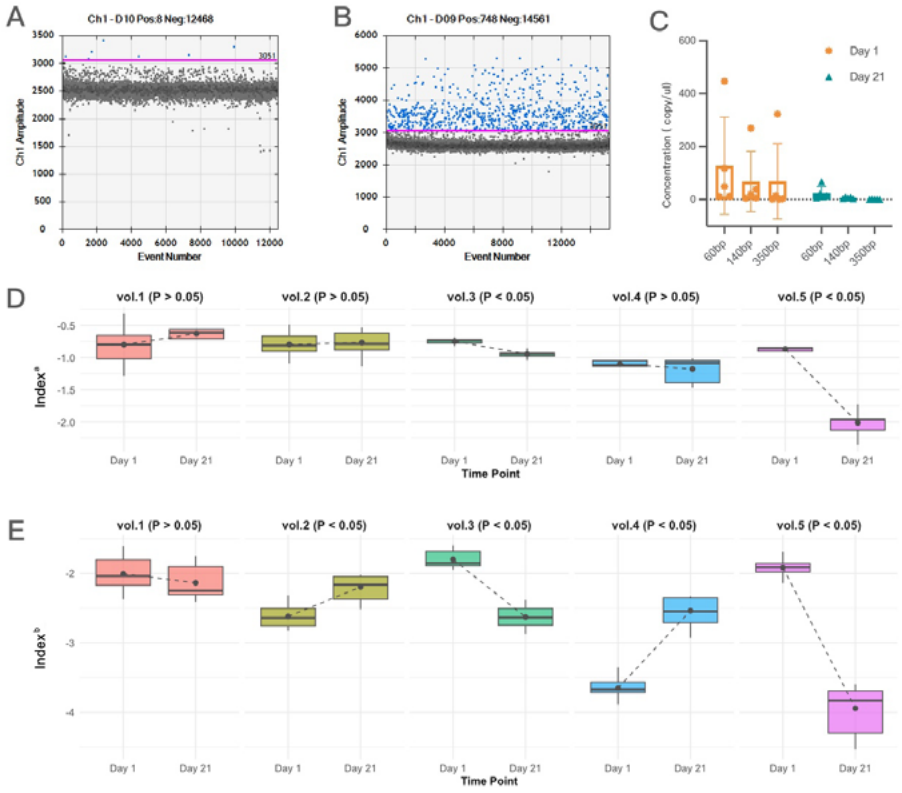


Figure 2. Absolute quantification results of standards based on digital PCR platform. The threshold fluorescence value for dividing negative and positive droplets is 3051. (A) Digital PCR results of negative blank control; (B) Digital PCR quantitative results of standards. (C) The copy number of fresh (0 day) and aged (21 days) fingerprints based on qPCR quantitative results. (D) Degradation trend chart of fingerprint samples based on degradation Index[a]. (E) Degradation trend chart of fingerprint samples based on degradation Index[b].

4. Discussion

This study successfully screened three TsD mtDNA markers with different length gradients (COI_{60bp}, COI_{140bp} and COI_{350bp}). In addition to the absence of Numts, species specificity, and conservatism in the human population, TsD genetic markers also met a series of conditions for amplification detection. The amplification specificity of the TsD markers was demonstrated through single peak melting curves, with all three primer pairs achieving amplification efficiencies exceeding 90%, thus validating their potential as candidate molecular markers for fingerprint TsD inferences.

Quantitative assessments were performed using absolute quantification of peripheral blood DNA on the ddPCR platform, providing standards for subsequent qPCR quantification. The initial standard concentration was 1160 copies/μl. Prior research has utilized fluorescence dye methods to estimate the bloodstain TsD over a period of up to 14 years [6]. Given the stability of the mtDNA structure, we conducted a quantitative assessment of three mtDNA markers in fresh fingerprint samples and those after 21-day interval. Our analyses of DNA extracted from fingerprint samples revealed a decline in mitochondrial gene concentrations over time, with 21-day aged fingerprints exhibiting lower levels compared to fresh fingerprints, indicating possible degradation of mtDNA. Furthermore, degradation indexes showed significant declines in some samples after 21 days, with $Index^b$ reflecting a higher degradation rate than $Index^a$. This finding aligns with previous research suggesting that longer nucleic acid fragments are more prone to degradation than shorter ones [7]. Notably, the study acknowledged the challenges associated with using fingerprints as trace biological samples [8], where some samples did not follow expected degradation patterns, potentially due to interference from fingerprint lipids and other factors. To enhance the reproducibility of TsD inference results, it is crucial to delineate the extent of various interferences and incorporate additional mitochondrial markers of varying lengths to mitigate the randomness of degradation. Overall, this research highlights the significant potential of mtDNA genetic markers for analyzing complex trace biological samples like fingerprints.

5. Conclusion

This study developed a complete strategy for screening mtDNA TsD markers, identifying three COI markers with high amplification specificity and efficiency

for fingerprint TsD inference. By analyzing the degradation patterns of mtDNA in fingerprints, this study demonstrated a quantifiable degradation pattern that can be used to infer fingerprint TsD, showcasing the sensitivity and effectiveness of the fluorescent probe method. Due to the particularity of fingerprint samples, future studies should expand the sample size and include more TsD genetic markers to improve the accuracy and reliability of fingerprint TsD inference.

6. Acknowledgments

This study was funded by the Chongqing post-doctoral Science Foundation (CSTB2023NSCQ-BHX0161).

7. Conflict of interest statement

The authors declare that they have no known conflicts of financial interest or personal relationships that could have influenced the work reported in this paper.

8. References

[1] A. Girod, R. Ramotowski, S. Lambrechts, P. Misrielal, M. Aalders, C. Weyermann, Fingermark age determinations: Legal considerations, review of the literature and practical propositions, Forensic Science International 262 (2016) 212-226.

[2] E. Brunelle, M. Eldridge, J. Halámek, Determination of Time since Deposition of Fingerprints via Colorimetric Assays, ACS Omega 6(19) (2021) 12898-12903.

[3] A. Coppin, Investigation of lipid degradation pathways for age determination of latent fingerprints, California State University, East Bay2023.

[4] K. Frisch, K.L. Nielsen, S. Francese, MALDI MSI Separation of Same Donor's Fingermarks Based on Time of Deposition—A Proof-of-Concept Study, Molecules 28(6) (2023) 2763.

[5] C. Weyermann, A. Girod-Frais, Estimating the Age of Fingermarks: Relevance, Potential Approaches, and Perspectives, in: J. De Alcaraz-Fossoul (Ed.), Technologies for Fingermark Age Estimations: A Step Forward, Springer International Publishing, Cham, 2021, pp. 59-83.

[6] H. He, Q. Zhang, Q. Niu, Y. Li, Q. Sun, D. Zhao, Estimating bloodstain formation time by quantitative analysis of mtDNA degradation, Forensic Sci Int 339 (2022) 111411.

[7] L. Bhoyar, P. Mehar, K. Chavali, An overview of DNA degradation and its implications in forensic caseworks, Egyptian Journal of Forensic Sciences 14(1) (2024) 15.

[8] C. Chen, Y. Lin, H. Yu, Y. Xue, Q. Xu, L. Jiang, R. Zhu, A. Chen, Y. Yang, R. Xia, X. Zhang, Q. Yang, R. Tao, B. Zhu, C. Li, S. Zhang, Establishing an integrated pipeline for automatic and efficient detection of trace DNA encountered in forensic applications, Sci Justice 62(1) (2022) 50-59.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1129-1136
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Assessment of simultaneous DNA/RNA/Protein extraction from human remains

Maria J. Castagnola[1], Francisco Medina-Paz[1], Sara C. Zapico[1,2*]

[1]*Department of Chemistry and Environmental Sciences, New Jersey Institute of Technology, Newark, NJ, 07102 USA.*

[2]*Department of Anthropology and Laboratories of Analytical Biology, National Museum of Natural History, Smithsonian Institution, Washington, DC, 20560 USA.*

** Corresponding author at: New Jersey Institute of Technology, Department of Chemistry and Environmental Science, Newark, NJ, 07102, sc338@njit.edu; and Smithsonian Institution, National Museum of Natural History, Anthropology Department and Laboratories of Analytical Biology, Washington, DC, 20560, casado-zapicos@si.edu*

Abstract

DNA isolation from human remains is of utmost importance to be able to identify a victim. However, there are situations in which, apart from the STR profile, other parameters need to be estimated, like the time-since-death. Recent advances in this field point to "-omics" technologies to improve these estimates, particularly in skeletonized remains. Among them, transcriptomics and proteomics are emerging as promising techniques for this purpose. However, this requires the extraction of RNA and proteins from the remains along with the DNA, which is particularly challenging in this kind of sample exposed to different environmental conditions.

This study aimed to assess the efficiency of simultaneous DNA/RNA/protein extraction from teeth, the hardest tissues in the human body and most resistant to decomposition. Towards this goal, six teeth were processed, separating dentin and pulp and creating aliquots of 200 ng of dentin.

The Zymo Quick-DNA/RNA™ Microprep Plus Kit was used for the simultaneous DNA/RNA/protein extraction. Different incubation times on the digestion buffer and proteinase K were evaluated. DNA/RNA/protein concentrations were assessed through Qubit fluorescent quantification, and human-specific DNA quantification was carried out by applying the Promega PowerQuant® kit.

According to our results, both Qubit and PowerQuant® quantification showed an increase in DNA yield after 24 hours of incubation, although not significant, obtaining average concentra-

tions around 30 ng/ul. Respective to RNA, time did not impact the RNA yield, obtaining average concentrations around 50 ng/ul. In contrast, for proteins, 2 hours of incubation increased the protein yield, obtaining average concentrations around 1 ug/ul. To the best of our knowledge, this is the first study assessing the efficiency of simultaneous DNA/RNA/proteins from human remains, demonstrating that it is possible to isolate these three molecules, obtaining appropriate yields, not only for human identification but also for time-since-death estimation.

Keywords
Simultaneous Extraction, Teeth, Human identification, Time-Since-Death Estimation

Introduction

DNA isolation is essential for identifying human remains during major disasters, humanitarian crises, and criminal investigations. However, environmental conditions such as humidity, temperature, and microbial activity can hinder the recovery of genetic material, often leaving skeletal remains—particularly teeth and bones—as the only viable sources of DNA.

Teeth are particularly valuable for forensic DNA extraction because of their hardness and resistance to environmental degradation. They consist of four distinct tissues: enamel, dentin, cementum, and pulp. The highly mineralized structure of enamel protects the nucleic acid-rich dentin and pulp from external conditions, making these inner tissues important targets for studies focused on human identification [1]. Multiple DNA extraction methods have been studied over time, with silica column and magnetic bead techniques being the most prevalent [2-4]. Despite this, there is no consensus on the best strategy.

In addition to generating DNA profiles, determining the postmortem interval (PMI) is critical for forensic investigations. Advances in "-omics" technologies, including transcriptomics and proteomics, are improving PMI estimation for skeletonized remains [5]. Moreover, the instability of RNA compared to DNA, highlights the need for optimized extraction methods. Selecting a technique that allows the simultaneous extraction of high-quality and high-quantity DNA, RNA, and proteins from tissue adds a layer of complexity, highlighting the necessity for optimized extraction methods.

This study focuses on optimizing the simultaneous extraction of DNA, RNA, and proteins from the dentin and pulp of teeth, as this initial stage is key for improving subsequent steps in human identification and estimating the postmortem interval (PMI).

Material and methods

Extraction

Six healthy, erupted third molars from five living donors, extracted between 5 and 10 years ago, were utilized in this study. Each tooth was irradiated for 15 minutes using UV light at 254 nm. Subsequently, a diamond cutting disk was employed to remove the cementum and enamel and to separate the crown from the root. The pulp was then separated from the dentin using tweezers. An agate mortar and pestle were used to reduce the dentin into thin fragments. Each sample was weighed and transferred to 1.5 mL Eppendorf tubes. Specifically, the dentin samples were divided into 200 mg aliquots in each tube. From each tooth, a single pulp sample and several dentin samples were obtained according to the dental mass.

DNA, RNA, and protein were extracted using the Quick-DNA/RNA Miniprep Plus Kit (Zymo Research). To each sample, 600 µL of DNA/RNA Shield, 30 µL of Proteinase K, and 60 µL of PK Digestion Buffer were added. The samples were incubated in a thermal mixer (Thermo Scientific) at room temperature with continuous agitation for different time points: 2 hours, 5 hours, 12 hours, and 24 hours.

After incubation, the tubes were vortexed, and the supernatant was transferred to new 1.5 mL tubes. The extraction process was then continued according to the kit's protocol. Final elution was performed using 15 µL of DNase/RNase-Free Water for DNA and RNA, and 200 µL of elution buffer for proteins.

Quantification

Quantification of DNA, RNA, and proteins was performed using different methods. Initially, the Nanodrop™ One Microvolume UV-Vis Spectrophotometer was performed to determine the presence of DNA/RNA and proteins in the samples. Additionally, the presence of salts in the extracts was assessed, and for DNA and RNA samples, potential contamination with proteins was also evaluated. In a second step, the Qubit™ Fluorometer (Thermo Scientific) was employed for more precise quantification. Finally, for DNA, human-specific quantification was performed using PCR with the Promega PowerQuant® Kit on the PowerQuant System (Promega Corporation), following the manufacturer's protocol. This analysis allowed for the evaluation of DNA quality by determining the degradation index (DI) and detecting potential inhibitors through the internal PCR control (IPC).

STR Amplification and Analysis

Nuclear DNA profiling was conducted by amplifying 23 autosomal STRs and the amelogenin gene using 1 ng of DNA per sample. This was performed with the

Promega PowerPlex® Fusion 6C System (Promega Corporation) on the ProFlex™ PCR System (Thermo Fisher Scientific) in a total volume of 25 µL, over 29 cycles. Subsequently, fragment analysis was carried out using the SeqStudio Genetic Analyzer (Thermo Scientific) with the following parameters: 7-second injection time, 1200 V injection voltage, 1440-second run time, and 9000 V run voltage. For the final analysis of DNA profiles, the Microsatellite Analysis Software in Thermo Fisher Cloud was used with a threshold of RFU 150 for peak detection.

Statistical analyses

Statistical analyses were carried out using SPSS® (IBM), including both parametric and non-parametric tests.

Results

DNA/RNA/protein quantification

Initially, the total nucleic acids and proteins were measured in all samples using Nanodrop™. The average values for dentin and pulp combined were 46.11 ng/µl for total DNA, 107.18 ng/µl for RNA, and 4.17 µg/µl for proteins. Following this, the Qubit fluorescence method was used for more precise quantification, resulting in values of 33.23 ng/µl for DNA, 57.17 ng/µl for RNA, and 1.08 µg/µl for proteins.

Specifically for human DNA quantification using PowerQuant®, the total DNA concentration was 21.68 ng/µl. The average concentration for dentin samples using this last method was 15.73 ng/µl, while pulp samples showed a concentration of 45.50 ng/µl. In terms of tissue differences, DNA yield was significantly higher in pulp compared to dentin (One-way ANOVA, p=0.009).

Due to the higher number of samples extracted from dentin (n=24) compared to pulp (n=6) and considering that all dentin samples had the same weight, the following analysis at different incubation times (2, 5, 12, and 24 hours) is presented exclusively for dentin samples.

The Qubit quantification specifically for dentin samples across all time intervals yielded the following results: DNA concentration 31.58 ng/µl (SD = 12.74), RNA concentration 51.42 ng/µl (SD = 33.04), and protein concentration 1.02 µg/µl (SD = 0.23). The data indicated no significant differences in DNA concentration (Kruskal-Wallis, p = 0.181) or RNA concentration (One-way ANOVA, p = 0.813) across different incubation times, with the highest levels observed at 24 hours.

Similarly, protein concentrations did not show significant variations (One-way ANOVA, p = 0.412) between the time intervals (Figure 1).

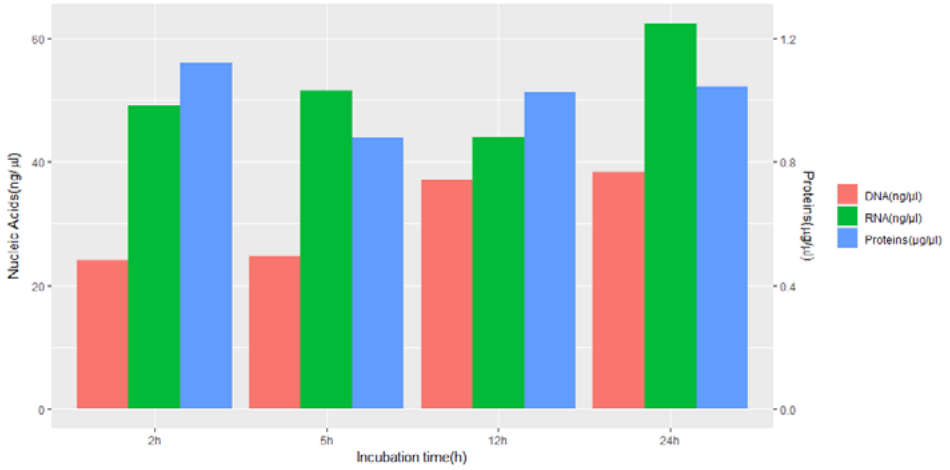


Figure 1. Concentration of nucleic acids (ng/µl) and proteins (µg/µl) as a function of incubation time.

Human DNA quantification using PowerQuant® showed that DNA yield increased with longer incubation times (24 hours), although this increase was not statistically significant (One-way ANOVA, p=0.064). The greatest deviation was observed at the 2-hour incubation time. Regarding the Quality Assessment, no sample inhibition was detected below the IPC shift. The degradation index values were below the threshold for all incubation times and samples, except for those from tooth 2, which showed a higher-than-expected degradation index (DI) ratio. This difference was statistically significant compared to all other teeth (ONE-WAY ANOVA, $p < 0.05$).

STR analysis

STR profiles were evaluated using the Promega PowerPlex® Fusion 6C System (Promega Corporation) on different samples. Dentin samples from tooth 3 were analyzed at various time intervals (2, 5, 12, and 24 hours), whereas pulp samples from multiple teeth (teeth 1–4) were studied at the same time interval (24 hours). All samples yielded complete DNA profiles.

Discussion

High quantities of DNA, RNA, and proteins were successfully extracted and quantified using different methods across all time points. Given their low porosity, high mineralization, and protection within the jaw, these results suggest that teeth are optimal samples for analyzing cadaveric material [6].

High-quality DNA was also demonstrated through human DNA quantification using PowerQuant®. For future analyses, RNA and protein quality could be assessed through gel electrophoresis. Additionally, proteins could be further analyzed using mass spectrometry, while specific RNA analysis could be investigated through RT-qPCR (Reverse Transcription-Quantitative polymerase chain reaction).

According to the data, both the Qubit and PowerQuant® quantification methods demonstrated an increase in DNA yield after 24 hours of incubation. Although this difference was not statistically significant, it suggests potential optimization of the extraction protocol by extending the incubation time to 24 hours for teeth. Despite the inherent concentration variations associated with each method, a consistent pattern was observed across different incubation times. Furthermore, the high standard deviation in the quantification results can be attributed to the nature of the sample [2, 7].

Specifically for quantification using PowerQuant®, no PCR inhibition was observed, which is an important factor for selecting the most efficient DNA extraction procedure [3]. In dentin samples, the increased degradation index was observed only in tooth 2, indicating also that this may be more related to the characteristics of the sample itself rather than the extraction efficiency.

Moreover, the pulp exhibited a significantly higher DNA yield compared to dentin ($p=0.009$), consistent with findings from previous studies [4]. This is expected, as the pulp is a highly vascularized tissue that contains multiple cell types [2, 8]. However, one notable advantage of dentin over pulp is the larger amount of sample available for extraction. On average, five to seven 200 mg Eppendorf tubes of dentin were obtained per tooth, compared to a single sample from the pulp. Therefore, both factors should be carefully considered when selecting tissue for analysis.

The successful amplification of complete DNA profiles across all incubation times (2, 5, 12, and 24 hours) demonstrates the robustness of the extraction method. The 24-hour samples, which yielded higher DNA concentrations and complete profiles, may offer advantages when multiple determinations are required from the same sample. Increased DNA yield could enable both individual identification and postmortem interval (PMI) estimation. Additionally, minimizing DNA loss is

crucial if DNA methylation analysis is carried out to improve age estimates, as the initial bisulfite treatment requires significant DNA quantities

Overall, these results show the importance of considering both tissue type and incubation time in optimizing DNA/RNA and protein extraction protocols. The high variability found among the samples highlights the need for future investigations to consider larger sample sizes and study teeth under various conditions—such as exposure to burning or acidic/alkaline soils—to better simulate real-life forensic scenarios.

Conclusion

In this study, high quantities of DNA, RNA, and proteins were simultaneously extracted from teeth using the Quick-DNA/RNA Miniprep Plus Kit (Zymo Research). Additionally, high-quality DNA samples were demonstrated through STR analysis. This method could be particularly advantageous for real case samples, maximizing the biological material available from DNA profiling to postmortem interval (PMI) estimation. To our knowledge, this is the first study to successfully demonstrate the simultaneous extraction of DNA, RNA, and proteins from teeth, providing important insights for DNA identification and time since death estimation.

Acknowledgments

María Josefina Castagnola is funded by the Fulbright Foreign Student Program – Ministry of Education Argentina.

Conflict of interest statement

All authors declare that they have no conflicts of interest.

References

[1] Barbaro A, Samar S, Falcone G, et al. Highly efficient and automated extraction of DNA from human remains using a modified EZ1 protocol. Forensic Sci Res. 2021;6(1):59-66.

[2] Zapico SC, Ubelaker DH. Sex determination from dentin and pulp in a medico-legal context. J Am Dent Assoc. 2013;144(12):1379-85.

[3] Barbaro A, Cormaci P, Falcone G. Validation of BTA™ lysis buffer for DNA extraction from challenged forensic samples. Forensic Science International: Genetics Supplement Series. 2011;3(1):e61-e2.

[4] Cafiero C, Re A, Stigliano E,et al. Optimization of DNA extraction from dental remains. Electrophoresis. 2019;40(14):1820-3.

[5] Procopio N, Bonicelli A. From flesh to bones: Multi-omics approaches in forensic science. Proteomics. 2024;24(12-13):e2200335.

[6] Chierto E, Cena G, Mann RW, et al. Sweet tooth: DNA profiling of a cranium from an isolated retained root fragment. J Forensic Sci. 2021;66(5):1973-9.

[7] Mann AE, Sabin S, Ziesemer K,et al. Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. Sci Rep. 2018;8(1):9822.

[8] Higgins D, Austin JJ. Teeth as a source of DNA for forensic identification of human remains: a review. Sci Justice. 2013;53(4):433-41.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1137-1142
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Assessment of DNA and mRNA stability under different conditions

David Moore[1*], Zuhaib Subhani[1], Jim Thomson[1]

[1]*Eurofins Forensic Services, Feltham, UK*

** Corresponding author at: Eurofins Forensic Services, david.moore@forensicsuk.Eurofins.com.*

Abstract

The stability DNA and mRNA are important to know under different conditions. Saliva was examined under different conditions. Saliva was stored dried on glass, fabric, and cotton swabs alongside liquid saliva at ambient temperature. DNA and mRNA yields were tested at time points between zero hours and up to 6 months.

Saliva was also dried on a few additional non-absorbent surfaces; glass beads, plastic counters and metal discs. Samples were tested at a number of different time periods from zero hours and 28 days.

The impact of storage temperature on the stability of DNA and mRNA was assessed with vaginal swabs and saliva swabs. Storage at ambient temperature, 4°C and -20°C were tested over a period of up to six months.

Samples were extracted and DNA was quantified with the PowerQuant® (Promega) system. mRNA recovery was assessed using an RT-qPCR assay targeting the FDCSP marker for saliva and the CYP2B7P1 marker for vaginal material.

The data showed that mRNA transcripts are generally stable, but storage on certain sample matrices (specifically hard, non-absorbent surfaces in the case of the FDCSP transcript) can lead to rapid degradation of mRNA, whilst DNA remains stable. Furthermore, DNA in liquid saliva can degrade with little to no impact on mRNA.

Keywords

Body fluid ID, mRNA, stability, sample matrix, temperature.

1. Introduction

DNA and mRNA are important targets used in the assessment of forensic evidence at sub-source and source levels. Knowing the stability and how different surfaces

impact the ability to detect these nucleic acids is important to better understand the likely results that might be observable from crime stain samples.

Here we look at the impact of different sample matrices on the recovery and detection of DNA and mRNA from saliva. In addition, to measuring the impact on DNA and mRNA stability from saliva and vaginal swabs stored under different temperature conditions.

2. Material studied, methods, techniques

2.1. Storage of saliva on different surfaces and storage of vaginal and saliva swabs under different temperature conditions.

Saliva was taken from three consenting donors. Several different surfaces/conditions were tested: (1) Dried on absorbent materials (cotton and cotton swabs) (2) Dried on non-absorbent material (Glass slides, stainless steel, polypropylene discs, glass beads). (3) stored as an aliquot of liquid saliva.

20µL of saliva from each donor was deposited for each condition. Where required, samples were allowed to dry for one hour (i.e. all samples except the liquid saliva aliquots). Twelve replicates were created for each donor/matrix combination. All samples were stored at ambient temperature (~20°C) and one of each replicate was tested at the timepoints: 0 hours (h), 1h, 6h, 24h, 48h, 7 days (d), 14d, 1 month (m), 3m, 6m, 12m.

Fabric swatches and liquid saliva samples were lysed directly in a 1.5mL Eppendorf tube. Non-absorbent surfaces were sampled with a cotton swab moistened with 20µL of water.

An additional set of saliva swabs and vaginal swabs (not dried) was stored at three different temperatures: ambient temperature (~20°C), 4°C and -20°C for up to 3 months and tested in two to three replicates at each time point.

Swab heads were placed into a 1.5mL Eppendorf tube prior to storage to allow for immediate lysis following storage.

2.2. Sample lysis, purification, quantification and body fluid identification processing

Lysis and purification used the Qiagen Investigator STAR Lyse&Prep kit modified for improved total nucleic acid purification. All samples were quantified using a reduced volume PowerQuant® assay.

mRNA expression was measured with a multiplex one-step RT-qPCR assay containing the following targets:

— CYP2B7P1 [1] (vaginal secretions, Hs00933441_g1 (FAM), Thermo Fisher))
— FDCSP [2] (saliva, Hs00395133_m1 (VIC), Thermo Fisher)).

These targets were chosen based on high expression within the body fluids of interest, and absence of false positives associated with expression in non-target body fluids and/or from the detection of human gDNA.

The assays both amplify small amplicons to improve detection of degraded mRNA targets. Assays were chosen that contained probes that span Exon-Exon junctions to provide specificity to the mRNA target over gDNA. An exogenous IPC (Quasar 670) was also included in the assay (Eurogentec). The GoTaq® Probe 1-Step RT-qPCR mix (Promega) was used in the assay with the fast-cycling conditions detailed within the user manual.

3. Results

3.1. Storage of saliva on different surfaces

Samples stored on absorbent surfaces such as cotton swatches and cotton swabs showed no notable change in DNA concentration or mRNA detection with regards to mRNA Cq value (Figure 1).

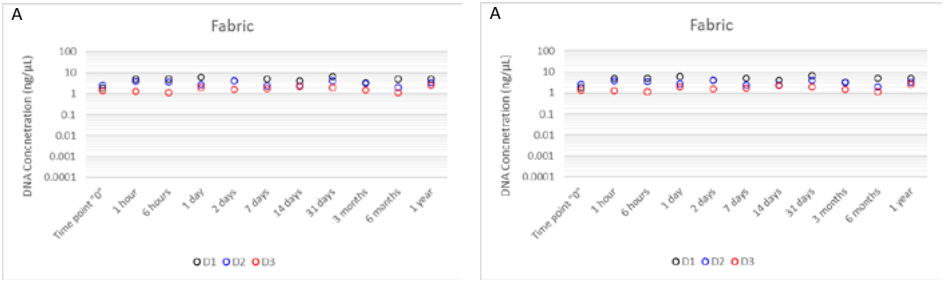


Figure 1. DNA yield (A) and mRNA detection (B) for saliva samples stored on fabric for up to one year. Donors are colour coded.

The samples stored as an aliquot of liquid saliva showed that DNA yield steadily declined over time. The decrease in DNA yield was clearly observable from seven

days. At six months, DNA yield had decreased two to four orders of magnitude. mRNA detection also decreased with Cq values steadily increasing up to until three months at which point some samples would not provide detectable results. After six months there were no detectable results. (Figure 2).

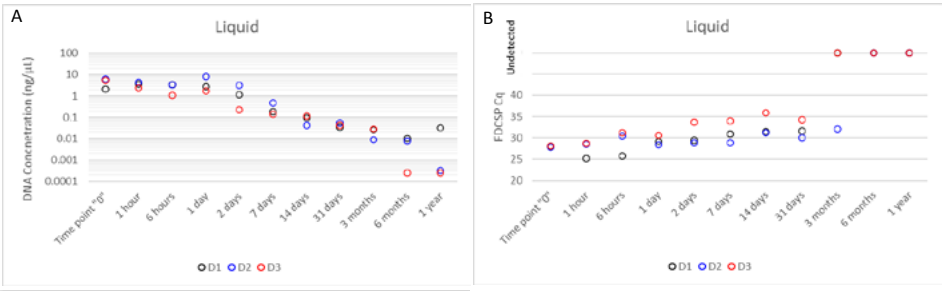


Figure 2. DNA yield (A) and mRNA detection (B) for saliva samples stored as liquid for up to one year. Donors are colour coded.

The samples stored on non-absorbent surfaces showed no notable change in DNA yield. However, there was a rapid reduction in mRNA detection. One of the glass samples showed no detection of the mRNA target even after a single hour of storage. There was no mRNA detected in any of the glass samples at one month.

This was a consistent effect observed across all hard surfaces tested (only glass slide results shown in Fig 3), where DNA showed no notable reduction in DNA yield, yet the mRNA component of the sample samples would rapidly degrade.

There seemed to be some association with donor in this study, with some donors failing to provide detectable levels of mRNA sooner than others.

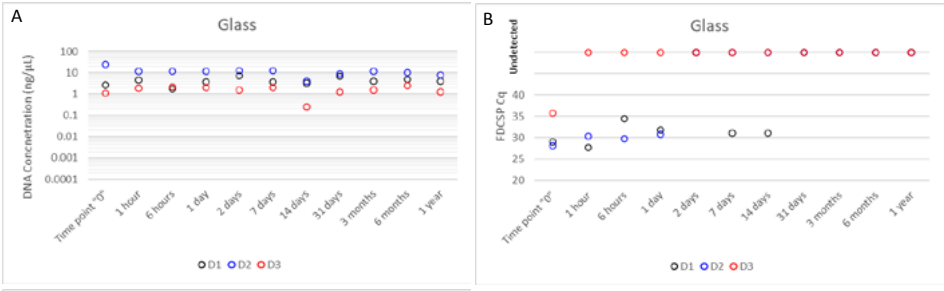


Figure 3. DNA yield (A) and mRNA detection (B) for saliva samples stored as dried stains on

glass for up to one year. Donors are colour coded.

3.2. Storage of saliva and vaginal material at different temperatures for up to 3m (ambient, 4°C,-20°C)

Saliva swabs stored undried at -20°C over a three-month period showed no notable change in DNA yield or change to mRNA Cq value. At 4°C there was no notable change in DNA yield over the three-month period. However, the mRNA detection was much more variable with some samples being undetected after as little as 24 hours and others being detected at all time points up to three months. The undried swabs stored at ambient temperature showed a notable decrease in DNA yield over time. The mRNA data showed rapid decreases in mRNA detection with mRNA not being detected all except one sample at or after 24 hours of storage, there was a single detectable observation at 48 hours.

Vaginal swabs stored undried at all temperatures (-20°C, 4°C and ambient) showed no notable change in DNA yield or change to Cq value over the entire three-month period of this study. Data are not shown for this experiment. These data suggest that there may be difference in sample stability between different body fluids, this may be associated with the mRNA molecules themselves being stable, or differences at the cellular level that may or may not be more detrimental to mRNA stability (data not shown).

4. Conclusion

The results from this investigation show that different sample matrices can impact the detection of mRNA over time that are not directly proportional to any impact on DNA. This is particularly true of non-absorbent surfaces where the ability to detect dried saliva was rapidly reduced. This observation appears to be true of several non-absorbent surfaces.

Samples that were seeded with saliva and left undried degraded much more rapidly than those that were dried, this indicates that the presence of moisture within a sample can hasten the loss of DNA and mRNA detection.

The CY2B7P1 mRNA target in vaginal material appears to be markedly more stable than that of the FDCSP marker in saliva. Further work is required to determine the impact of non-absorbent surfaces of mRNA on vaginal material. Additionally, studies looking at whether these body fluid related differences are entirely related to body fluid or whether gene target of mRNA target location (i.e. different

exon-exon junctions) have an impact on observed stability, as has previously been suggested [3]

5. Acknowledgments

The authors are grateful to all volunteers who donated samples for this work

6. Conflict of interest statement

The authors declare they have no conflicts of interest

7. References

[1] RNA/DNA co-analysis from human menstrual blood and vaginal secretion stains: Results of a fourth and fifth collaborative EDNAP exercise. Haas, C. et al. Forensic Science International: Genetics, Volume 8, Issue 1, 203 – 212

[2] Genome-wide mRNA profiling and multiplex quantitative RT-PCR for forensic body fluid identification Park, Seong-Min et al. Forensic Science International: Genetics, Volume 7, Issue 1, 143 - 150

[3] Degraded RNA transcript stable regions (StaRs) as targets for enhanced forensic RNA body fluid identification Lin, Meng-Han et al. Forensic Science International: Genetics, Volume 20, 61 - 70

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1143-1149
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Comparative Evaluation of Targeted RNA Sequencing Protocols for Gene Expression Quantification With and Without Unique Molecular Indices (UMIs)

Annica Gosch[1], Cornelius Courts[2*]

[1]*Institute of Legal Medicine, University Hospital Cologne, Cologne, Germany*

** Corresponding author at: Melatengürtel 60-62, 50823 Cologne, Germany, cornelius.courts@uk-koeln.de*

Abstract

Interest in forensic RNA analysis has increased over the last years. RNA molecules present in forensic samples can accurately be quantified via quantitative PCR (qPCR), however, due to the limited number of markers that can be assayed simultaneously per reaction, qPCR is less suitable for applications requiring gene expression quantification of large marker sets. Few years ago, massively parallel targeted RNA-sequencing (targRNAseq) allowing to simultaneously and accurately quantify several hundreds of markers has been added to the forensic genetic tool set. However, typical targRNAseq protocols include a multiplex-PCR-step to amplify selected targets which potentially introduces bias and limits accurate gene expression quantification.

Unique Molecular Indices (UMIs) have been invented to overcome this limitation and have been implemented in protocols from some vendors.

In this study, we compared two targeted RNAseq protocols assaying expression of a set of 121 forensically relevant mRNA biomarkers: The Ion Ampliseq targeted RNA sequencing panel (Thermo Fisher Scientific), which employs a multiplex-PCR without the use of UMIs, and the QIAseq targeted RNA panel (QIAGEN), which uses UMIs prior to multiplex amplification.

Both protocols were tested on replicated samples and dilution series and compared with respect to sensitivity and accuracy of gene expression quantification.

The UMI-based protocol exhibited decreased sensitivity in comparison to the non-UMI-based alternative, however, making use of UMI technology greatly improved gene expression quantification accuracy. We thus recommend the use of UMI-based protocols for targeted RNA sequencing for applications requiring accurate gene expression quantification.

Keywords

forensic RNA analysis, targeted RNA sequencing, gene expression quantification, unique molecular indices

1. Introduction

Analyses of differential gene expression to identify body fluids and/or organ tissues present in biological stains are routinely performed by several forensic laboratories worldwide and other potential applications of forensic RNA analysis are currently being explored.

Gene expression can accurately be quantified via quantitative PCR (qPCR) however, this method is less suitable for applications requiring the simultaneous analysis of large marker sets. Few years ago, massively parallel targeted RNA sequencing (targRNAseq) enabling simultaneous quantification of several hundreds of markers has been adopted for forensic molecular analyses. However, typical targRNAseq protocols include a multiplex-PCR to amplify selected targets which potentially introduces bias and limits accurate gene expression quantification. Unique Molecular Indices (UMIs), barcodes tagged to individual target molecules prior to PCR amplification and thus allowing to differentiate whether sequencing reads have been generated from a single or several distinct original target molecules, have been devised to overcome this limitation and have since been implemented in workflows and kits from some vendors.

Here, we assess whether the utilization of UMIs provides an advantage in targRNAseq for gene expression quantification in forensic samples. To this purpose, we compared two targRNAseq protocols to quantify the expression of a set of 121 forensically relevant mRNA biomarkers: The Ion AmpliSeq Library preparation kit (Thermo Fisher Scientific), in which multiplex PCR is performed without the use of UMIs, and the QIAseq targeted RNA panel (QIAGEN), which introduces UMIs prior to multiplex amplification.

2. Material studied, methods, techniques

2.1 Samples

Blood samples were obtained by fingerprick from four volunteers (2 males, 2 females) at two different time points of the day (donors A and B) or at a single time point (donors C and D). Samples were stored at -80°C immediately (donors A and B) or after a drying period at room temperature for 2 days (donor D) or for 2 days and 16 days (donor C, two samples taken directly after each other, stored for different time periods). Total RNA was extracted using the mirVana miRNA isolation kit (Thermo Fisher Scientific) following the manufacturer's instructions

including a lysis step for 30 min at 56°C. Genomic DNA was removed from the RNA extracts using the TURBO DNA free kit (Thermo Fisher Scientific) according to the manufacturer's protocol. RNA concentrations were then quantified using the Qubit High Sensitivity Assay on a Qubit 2.0 Instrument (both Thermo Fisher Scientific).

2.2 Library Preparation and Sequencing

Sequencing libraries were prepared using the Precision ID Library kit with Precision ID IonCode™ Barcode Adapters (Thermo Fisher Scientific) and the QIAseq targeted RNA panel with the QIAseq Targeted RNA 96-Index HT L for Ion Torrent (QIAGEN) according to manufacturer's instructions ("Ion AmpliSeq RNA libraries" protocol from the Ion AmpliSeq Library Kit Plus User Guide and "Ultra-low input and FFPE Sample v2" protocol, respectively). Custom primer panels targeting 121 forensically relevant mRNA markers for body fluid identification [1] and time of day estimation [2] were designed by and obtained from each manufacturer separately.

For comparison of gene expression accuracy, sequencing libraries were prepared in triplicates (for two samples from donor A) or duplicates (for two samples from donor B) with an input amount of 20 ng total RNA per library.

For assessing the assay sensitivity, sequencing libraries were prepared in a three-point dilution series with input amounts of 20 ng, 5 ng and 0.5 ng (for three samples from donors C and D).

Libraries were quantified by qPCR (Ion Library TaqMan™ Quantitation Kit (Thermo Fisher Scientific) and QIAseq™ Library Quant Assay Kit (QIAGEN), respectively) and equimolarly pooled for sequencing on two Ion 530 sequencing chips. Templating and sequencing were performed using the Ion S5™ Precision ID Chef & Sequencing Kit on an Ion Chef and an Ion GeneStudio S5 system instrument, respectively (all Thermo Fisher Scientific).

2.3 Data analysis

Raw sequencing data analysis was performed on the Ion Torrent Suite platform (Thermo Fisher Scientific). For the Ion AmpliSeq-libraries, raw read counts were obtained as end-to-end-reads from sequencing data using the CoverageAnalysis plugin from the Ion Torrent Suite (Thermo Fisher Scientific). For the QIAseq libraries, raw read counts and raw UMI counts were obtained using the QIAseq targeted RNA panels gene expression pipeline on the web-based MyGeneGlobe platform (QIAGEN).

Raw read counts/raw UMI counts were rlog-transformed and Principal Component Analysis (PCA) was performed using the DEseq2 [3] package (version 1.42.1) in R (version 4.3.2).

3. Results

3.1 Accuracy

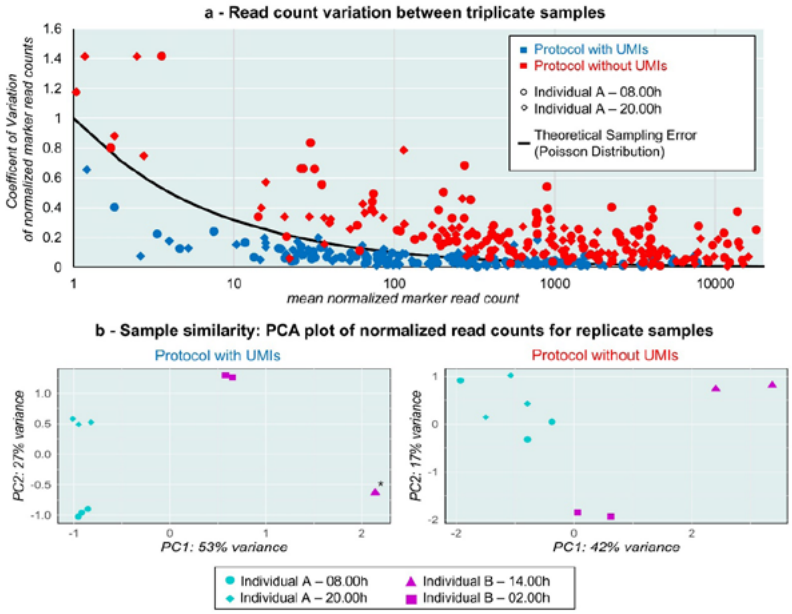


Figure 1. Comparison of gene expression quantification accuracy between sequencing libraries prepared with and without UMIs

a – Coefficient of variation between normalized marker read counts for two samples analyzed in triplicates with each protocol.

b – Principal Component Analysis for samples from two individuals taken at two different time points each, analyzed in triplicates (donor A) or duplicates (donor B). *Data points for individual B (14.00h) are directly above each other so that they cannot be visually separated.

Replicate samples analyzed with the UMI-based protocol exhibited lower coefficients of variation than replicate samples from the protocol without UMIs

(Figure 1a). Accordingly, in the PCA plot a tighter clustering of replicate samples and a better separation of distinct samples can be observed for the protocol with UMIs (Figure 1b).

3.2 Sensitivity

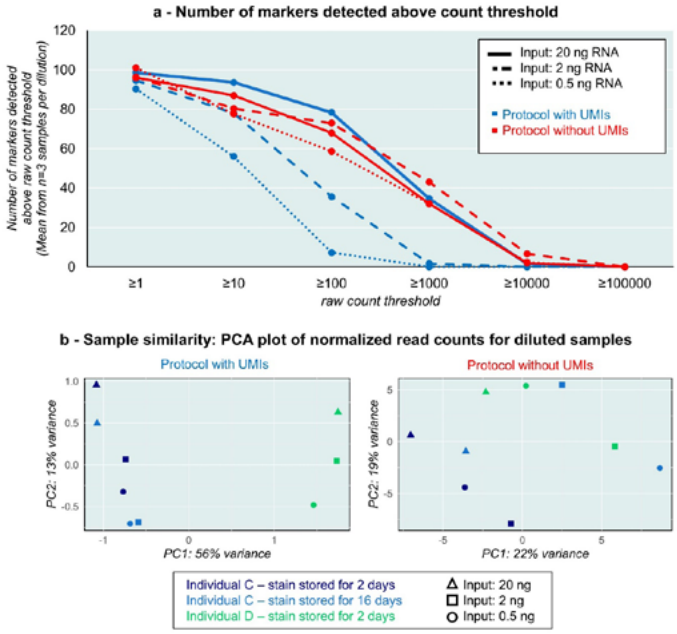


Figure 2. Comparison of gene expression quantification sensitivity between sequencing libraries prepared with and without UMIs

a – Number of markers detected above a raw read count threshold for sequencing libraries prepared from 20 ng, 5 ng and 0.5 ng total RNA with each protocol. Indicated values represent an average from three samples per dilution.

b – Principal Component Analysis for samples from two individuals dried for two different time periods at room temperature (2 days and 16 days), analyzed in at RNA input amounts of 20 ng, 2 ng and 0.5 ng.

For the UMI-based protocol, marker detection was decreased with lower sample input amounts while this was not observed for the protocol without UMIs (Figure 2a). However, samples sequenced with the UMI-based protocol showed better separation according to biological properties even for the diluted samples,

whereas samples from distinct donors were not separated in the PCA plot for the protocol without UMIs (Figure 2b).

4. Discussion

The UMI-based protocol resulted in more accurate gene expression quantification compared to the library preparation protocol without UMIs, as evidenced by lower technical variation. This is in line with the theoretically expected advantage of UMIs reducing PCR-induced bias.

Marker detection sensitivity was reduced in the UMI-based protocol, which may be caused by loss of sample material in the molecular tagging step or during the bead clean-up steps in the library preparation protocol. However, while markers were still detected from very low input amounts with the protocol without UMIs, quantitative information from these measurements has to be considered less reliable as evidenced by a loss of the ability to distinguish distinct samples in PCA.

5. Conclusion

Replicate samples analyzed with the UMI-based protocol exhibited a lower variation than replicate samples analyzed with the protocol without UMIs. The detection sensitivity was reduced in the protocol with UMIs compared to the protocol without UMIs, however, biological differences (inter-individual differences, differences between sampling time points) were better captured with the UMI-based protocol. We thus recommend the use of UMI-based protocols for targeted RNA sequencing for applications requiring accurate gene expression quantification.

6. Acknowledgments

We are grateful to the anonymous sample donors for their participation in this study. We thank the Deutsche Forschungsgemeinschaft (DFG) for funding this study (CO-992/10-2).

Conflict of interest statement

None.

References

[1] E. Hanson, G. Dørum, M. Zamborlin, S. Wang, M. Gysi, S. Ingold, R. Lagace, C. Roth, C. Haas, J. Ballantyne, Targeted S5 RNA sequencing assay for the identification and direct association of common body fluids with DNA donors in mixtures, Int. J. Legal Med. (2022). https://doi.org/10.1007/s00414-022-02908-9.

[2] A. Gosch, A. Bhardwaj, C. Courts, TrACES of Time: Transcriptomic Analyses for the Contextualization of Evidential Stains – identification of RNA markers for estimating time-of-day of bloodstain deposition, Forensic Science International: Genetics (2023) 102915. https://doi.org/10.1016/j.fsigen.2023.102915.

[3] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550. https://doi.org/10.1186/s13059-014-0550-8.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1151-1158
DOI: https://dx.doi.org/10.15304/cc.2025.1869

A nanopore sequencing-based mRNA set for body fluids and personal source analysis

Suyu Li[1], Mengyao Zhao[1], Jing Liu[1], Hong Zhu[1], Yiping Hou[1], Zheng Wang[1*]

[1] *Institute of Forensic Medicine, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu610041, China*

** Corresponding author at: No.24, South Section 1, Yihuan Road, Chengdu, Sichuan, China, wangzhengtim@scu.edu.cn, Tel: +86-28-85501549; Fax: +86-28-85501549.*

Abstract

Personal identification can determine the donor of biological traces left at crime scenes, and body fluids identification helps to link the donor to the criminal behavior, both of which are crucial to the investigation and prosecution of a case. Recent studies have shown that the differential expression patterns of mRNA can identify body fluids, and its sequence variation can be used to distinguish donors. Long fragment reading of nanopore sequencing can reduce the effects of PCR bias and fragment splicing comparing massively parallel sequencing (MPS) and find more base variation. In this study, we developed a targeted multiple nanopore mRNA sequencing system for forensically relevant body fluid identification (peripheral blood, menstrual blood, saliva, semen, and vaginal secretion), containing 12 mRNA markers (CD3G and CD93 for peripheral blood, MMP10, MMP7 and MMP11 for menstrual blood, PRB4 and STATH for saliva, SEMG1 and SEMG2 for semen, FAM83D and DKK4 for vaginal secretion, and LCE1C for skin) together with two reference genes (ACTB and GAPDH). The results demonstrated that in peripheral blood, semen and menstrual blood samples, only the specific mRNAs were highly expressed, with the average reads proportion of 99.38%, 99.99% and 97.59%, respectively. However, in saliva and vaginal secretion samples, due to the expression of vaginal secretion markers FAM83D and DKK4 in saliva and skin markers LCE1C in vaginal secretions, the average reads of specific mRNA accounted for only 66.60% and 64.49%, respectively. However, in saliva and vaginal secretion samples, the average read of specific mRNAs accounted for 66.60% and 64.49%, respectively. Despite the presence of cross-reactivity, each of the body fluid has unique mRNA profiles, allowing effective identification of body fluids. In addition, 41 cSNPs were contained in 14 mRNA transcripts, which helps to establish a link between the body fluids and the donors, but more cSNPs are needed to improve the power of discrimination.

Keywords
Forensic genetics, Body fluid identification, mRNA, Coding single nucleotide polymorphism (cSNP), Nanopore sequencing

1. Introduction

Identifying the body fluids and personal source of biological stain deposited at crime scenes can establish an important link between the donor and the criminal activity, and contribute to the investigation and prosecution of cases [1]. With advancements in genomics and technologies, body fluid identification has shifted from traditional protein-based methods to a variety of genetic markers, including DNA methylation, copy number variations, mRNA, microRNA, lncRNA, circRNA, piRNA, and microbiomes [2]. Notably, the tissue/body specificity of mRNA and its feasibility for co-analysis with DNA have been validated through several collaborative researches [3-6].

However, the independent detection and analysis of DNA-based personal identification and RNA-based body fluid may lead to "the association fallacy" [7]. Omedei et al. [8] proposed coding region SNPs (cSNPs) in body fluid-specific mRNA can assign body fluids to donors. Some research groups have developed different cSNP panels based on SNaPshot sequencing and massively parallel sequencing (MPS) technology platforms for identifying single type of body fluid donors and several types of body fluid contributors in mixed samples [9-11]. In addition, a collaborative exercise analyzed body fluid stains with a 33plex mRNA panel and a 35plex cSNPs panel to simultaneously identify body fluids types and donors [12]. Compared to MPS, Nanopore sequencing, with its ability to read long fragments, can reduce the effects of PCR bias and fragment splicing, and make it possible to detect more cSNPs in specific mRNAs using fewer primers, reducing the complexity of amplification system construction.

Therefore, in the present study, we developed a nanopore sequencing-based mRNA set comprising 12 specific mRNAs and 2 reference genes for the simultaneous analysis of body fluids and personal sources, with 41 cSNPs contained in the 14 mRNA transcripts.

2. Material studied, methods, techniques

2.1. mRNA selection and primer design

Specific mRNAs were selected from previous collaborative exercises [6, 13], with ACTB and GAPDH commonly used as reference genes (RGs) [14, 15]. Using

Primer-Blast, primers were designed to span at least one exon-exon junction to avoid genomic DNA contamination, and amplicon length was set between 500 bp and 1000 bp to minimize amplification bias. Primer specificity was validated through 1.5% agarose gel electrophoresis and nanopore sequencing. Details of mRNAs and primers used for multiplex PCR were shown in **Table 1**.

Table 1. Characteristics of candidate reference genes and body fluid-specific mRNAs.

Genes	Chr.	PCR primers	Production (bp)	cSNPs number	Specificity
ACTB	7	F: GGCTGTGCTATCCCTGTACG R: CTTGATCTTCATTGTGCTGGGT	574	1	RG
GAPDH	12	F: GGCAAATTCCATGGCACCG R: AAGTGGTCGTTGAGGGCAAT	761	2	RG
CD3G	11	F: CCGGAGGACAGAGACTGACA R: GGAGAACACCTGGACTACTCTG	600	3	PB
CD93	20	F: GATGAGTGTGCTCTGGGTCG R: CCCAGGTGTCGGACTGTACT	789	5	PB
PRB4	12	F: TGAGAGTTCAAGTGAAGATGTCAGC R: ACGGCATTTGTAGCAATTGAATTAT	771	5	SA
STATH	4	F: GTCTCATCTTGAGTAAAAGAGAACC R: ACCTCTGACATTATGGGCTTC	630	2	SA
SEMG1	20	F: AAACCTCACTCTGTCCTGCG R: ACTGAGGTCAACTGACACCTTG	748	5	SE
SEMG2	20	F: AACTGTGGAGAAAAGGGCATCC R: TGAGGTCGGGTGACACCTTG	735	2	SE
FAM83D	20	F: GCAACAGGCTCCTACAGTTTTAC R: GTTGAGCCGGGACCTGATAC	870	1	VA
DKK4	8	F: AGCCGATTTCACGCACCTTAC R: TTCTCCCTCTTGTCCCTTCCTG	602	1	VA
MMP10	11	F: ACCCACCTTACATACAGGATTGT R: CCTCTGATGGCCCAGAACTC	731	3	MB
MMP7	11	F: GTGGGAACAGGCTCAGGACTAT R: GAATGAATGGATGTTCTGCCTGA	740	3	MB
MMP11	22	F: GATCGACTTCGCCAGGTACT R: TTTCACCGTCGTACACCCAG	638	1	MB
LCE1C	1	F: GTGACCCCGCTCCTGAA R: GCACGCTGCAAATGACATTGA	646	7	SK

2.2. Samples preparation, multiplex PCR and nanopore sequencing

56 samples of peripheral blood (PB), saliva (SA), semen (SE), vaginal secretions (VA) and menstrual blood (MB) were collected from unrelated healthy individuals who had given informed consent. Total RNA was isolated, quantified, and then reverse transcribed into cDNA. Multiplex PCR was optimized by adjusting RNA input, primer concentration, annealing temperature, and cycle number. Library preparation was completed by end-repair, adapter ligation, purification and quantification, followed by nanopore sequencing of each library individually using the QNome-3841.

2.3. Data analysis

The sequencing data were filtered with Q-value > 8 and then aligned to reference sequences composed of the targeted cDNAs. The sequence data that met the following criteria—coverage > 90% and fewer than 40 bp of soft-clipped at the left and right ends—were analyzed expression patterns of 12 mRNAs in different body fluids based on the reads proportion. The SNPs detected by sequence were aligned to 1000 Genomes Phase 3 data, filtering for those with a minimum allele frequency (MAF) ≥0.1. Sequencing performance metrics, linkage disequilibrium of cSNPs and the power of discrimination (PD) for each body fluid were calculated.

3. Results

We developed a targeted multiplex nanopore mRNA sequencing assay for body fluid and personal source identification using the QNome-3841 platform. The multiplex PCR amplified 12 specific mRNA markers along with two reference genes (RGs), yielding amplicon lengths ranging from 574 to 870 bp. The amount of data among 56 samples ranged from 540 to 1,037 Mbp, with an average of 1,104,819 total reads and 1,103,260 mapping reads per sample. The median percentage of noise levels for all samples were below 1% with the exception of one MB sample (5%).

The reference genes ACTB and GAPDH were stably and highly expressed in 56 samples. Figure 1 showed the expression levels of 12 specific mRNAs in five body fluids, excluding the reference genes. We found that only CD3G and CD93 were expressed in PB, while SEMG1 and SEMG2 were exclusively expressed in SE. The average read proportions of PB- and SE-specific mRNAs in their target body fluids were 99.38% and 99.99%, respectively. Although in MB CD3G and CD93 were expressed in MB, the average read proportions of MMP7, MMP10, and MMP11

reached 97.59%. In SA, the expression level of PRB4 was significantly higher than that of STATH, with the average read proportion of 66.60% for these two SA-specific mRNAs. Additionally, other specific mRNAs were also expressed, with VA-specific mRNAs showing the highest average proportion at 20.62%. FAM83D and DKK4 were highly expressed in VA, and the average reads proportion of both was 64.49 %. However, high expression of the skin biomarker LCE1C was also observed in 8 out of 11 vaginal secretion samples (22.13%).

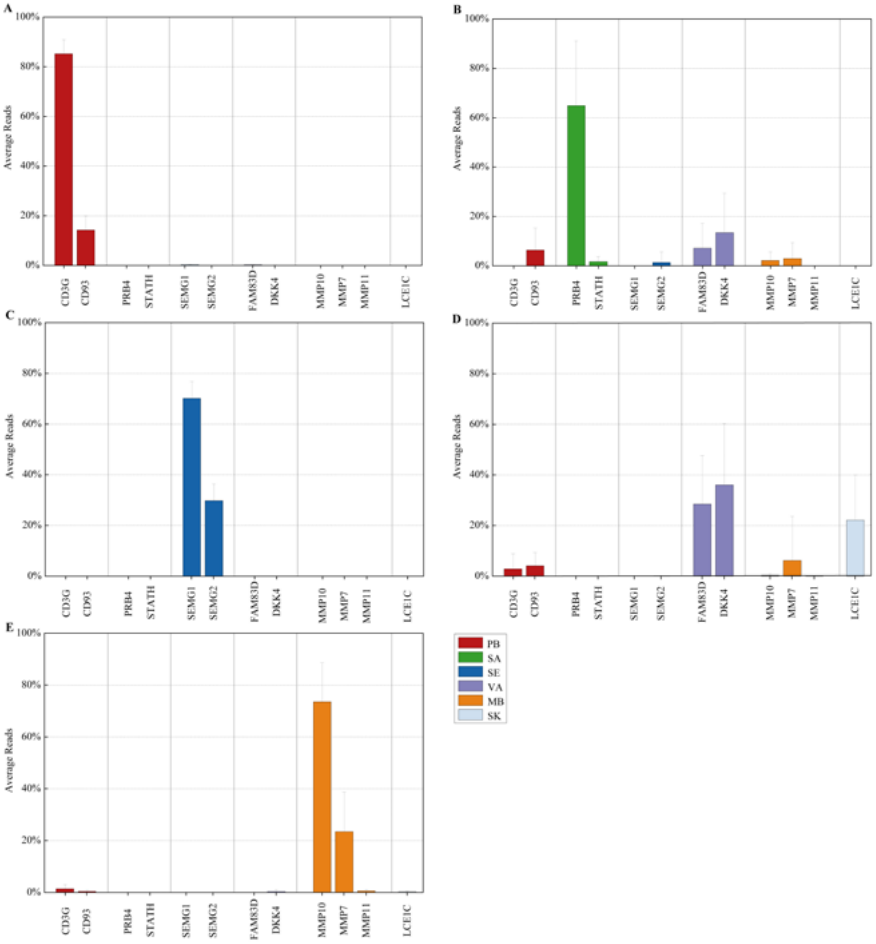


Figure 1. The Average reads proportion of 12 mRNAs in different body fluid samples. Different colors indicate the type of body fluid to which mRNA belongs. A, B, C, D and E correspond to PB, SA, SE, VA and MB samples, respectively.

According to 1000 Genomes Phase 3 data, 41 cSNPs with MAF≥0.1 were contained in 14 mRNAs, forming 37 RNA microhaplotypes (MHs). Using LDlink software, we evaluated the linkage between different genes on the same chromosome and found that SEMG1 and SEMG2 are linkage disequilibrium (LD). cSNPs within the same gene were considered genetically linked. We analyzed the phenotypic frequencies of each MH form 1000 Genomes Phase 3 data and calculated the total probability of discrimination power (TDP) for cSNPs within each body fluid. Given the high expression of ACTB and GAPDH in all body fluids, cSNPs of these reference genes were included in the TDP calculations. PB exhibited the highest TDP of 0.90, while SA and MB had TDPs of 0.78 and 0.79. SE and VA had lower TDPs of 0.66 and 0.56, respectively. However, LCE1C was highly expressed in most VA samples (8/11), suggesting it may serve as a supplementary marker for personal identification in VA (TDP 0.82). Due to our small sample size (56 samples), we only detected alternate alleles of 15 cSNPs with frequencies ranging from 4.39% to 77.98% and 15 RNA MHs.

4. Discussion

In the nanopore sequencing-based mRNA set, the average reads proportions of PB-, SE-, and MB-specific mRNAs in their respective target body fluids exceeded 95%, demonstrating high specificity. However, significant cross-reactivity was observed in SA and VA, particularly with the skin biomarker LCE1C, which was highly expressed in VA. This may result from contact with the outer surface of the skin during self-sampling. Despite the cross-reactivity, each body fluid exhibits unique mRNA profiles, allowing effective identification of body fluids. PB, SA, and MB exhibited strong persona identification ability, while identifying individuals from SE and VA presented more challenging. This may be because SEMG1 and SEMG2 in LD were merged into one genotype, reducing the number of markers. Additionally, FAM83D and DKK4 each contain only one cSNP. To improve the power of discrimination for body fluids and donors, we need to incorporate more highly expressed specific mRNAs, and design longer amplicons to include more cSNPs.

5. Conclusion

In this study, using QNome nanopore sequencing, we developed a mRNA detection system consisting of 12 body fluid-specific mRNAs and 2 RGs to identify the 5 body

fluids. In addition, we found 41 cSNPs in the mRNAs, and the TDPs for 5 body fluids were 0.56-0.90. This study demonstrates the potential of nanopore sequencing for simultaneously analyzing the sources of body fluids and donors. Further efforts will focus on expending mRNA sets and comprising larger samples to meet the requirements of forensic practice.

6. Acknowledgments

We would like to thank Liu Qin for her technical help. This study was supported by grants from the Major Program of National Nature Science Foundation of China (82293653). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

7. Conflict of interest statement

The authors declare that they have no conflict of interest.

8. References

[1] T. Sijen and S. Harbison, On the Identification of Body Fluids and Tissues: A Crucial Link in the Investigation and Solution of Crime, Genes 12 (11) (2021).

[2] T. Sijen, Molecular approaches for forensic cell type identification: On mRNA, miRNA, DNA methylation and microbial markers, Forensic Sci. Int. Genet. 18 (2015) 21-32.

[3] C. Haas, E. Hanson, M.J. Anjos, et al., RNA/DNA co-analysis from human menstrual blood and vaginal secretion stains: Results of a fourth and fifth collaborative EDNAP exercise, Forensic Sci. Int. Genet. 8 (1) (2014) 203-212.

[4] C. Haas, E. Hanson, M.J. Anjos, et al., RNA/DNA co-analysis from human saliva and semen stains – Results of a third collaborative EDNAP exercise, Forensic Sci. Int. Genet. 7 (2) (2013) 230-239.

[5] C. Haas, E. Hanson, M.J. Anjos, et al., RNA/DNA co-analysis from blood stains—Results of a second collaborative EDNAP exercise, F Forensic Sci. Int. Genet. 6 (1) (2012) 70-80.

[6] S. Ingold, G. Dørum, E. Hanson, et al., Body fluid identification using a targeted mRNA massively parallel sequencing approach – results of a EUROFORGEN/ EDNAP collaborative exercise, Forensic Sci. Int. Genet. 34 (2018) 105-115.

[7] P. Gill, Misleading DNA Evidence: Reasons for Miscarriages of Justice, first ed., Academic Press, London, 2014.

[8] M. Omedei, S. Gino, S. Inturri, et al., Individual assignment of body fluids in mixed stains by synonymous SNP analysis, Forensic Sci. Int. Genet. Suppl. Ser. 4 (1) (2013) e19-e20.

[9] E. Hanson, S. Ingold, G. Dorum, et al., Assigning forensic body fluids to DNA donors in mixed samples by targeted RNA/DNA deep seqeuncing of coding region SNPs using ion torrent technology, rensic Sci. Int. Genet. Suppl. Ser. 7 (1) (2019) 23-24.

[10] S. Ingold, G. Dorum, E. Hanson, et al., Assigning forensic body fluids to donors in mixed body fluids by targeted RNA/DNA deep sequencing of coding region SNPs, Int. J. Legal Med. 134 (2) (2020) 473-485.

[11] J. Liu, X. Cheng, F. Liu, et al., Identification of coding region SNPs from specific and sensitive mRNA biomarkers for the deconvolution of the semen donor in a body fluid mixture, Forensic Sci. Int. Genet. 52 (2021) 102483.

[12] S. Ingold, G. Dorum, E. Hanson, et al., Body fluid identification and assignment to donors using a targeted mRNA massively parallel sequencing approach - results of a second EUROFORGEN / EDNAP collaborative exercise, Forensic Sci. Int. Genet. 45 (2020) 102208.

[13] E. Hanson, S. Ingold, C. Haas, et al., Messenger RNA biomarker signatures for forensic body fluid identification revealed by targeted RNA sequencing, Forensic Sci. Int. Genet. 34 (2018) 206-221.

[14] A. Lindenbergh, M. de Pagter, G. Ramdayal, et al., A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces, Forensic Sci. Int. Genet. 6 (5) (2012) 565-577.

[15] M. Visser, D. Zubakov, K.N. Ballantyne, et al., mRNA-based skin identification for forensic applications, Int. J. Legal Med. 125(2) (2011) 253-263.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1159-1164
DOI: https://dx.doi.org/10.15304/cc.2025.1869

A sequencing assay incorporating coding region SNPs for body fluid identification

Tingting Jiang[a], Guanqing Ji[c], Ruocheng Xia[a], Shouyu Wang[b], Ruiyang Tao[a*]

[a]*Shanghai Key Laboratory of Forensic Medicine, Shanghai Forensic Service Platform, Academy of Forensic Sciences, Key Laboratory of Forensic Science, Ministry of Justice, Shanghai 200063, China.*

[b]*Department of Forensic Medicine, Shanghai Medical College, Fudan University, Shanghai 200032, China.*

[c]*Public Security Bureau of Tongzhou, Nantong 226300, China.*

** Corresponding author at: No. 1347 West Guangfu Road, Putuo District of Shanghai, China, taoruiyang163@163.com.*

Abstract

Objetives: Body fluid identification (BFID) and tracing the donors of specific components are essential aspects of forensic practice. In recent years, coding region single nucleotide polymorphisms (cSNPs) in body fluid-specific mRNAs have emerged as significant molecular markers for BFID, enabling direct association of specific body fluids with their respective donors. In this study, we conducted and applied cSNP analysis using a highly sensitive targeted RNA sequencing assay. **Material and methods:** In our recent work, we screened body fluid-specific mRNAs containing coding region single nucleotide polymorphisms (cSNPs) with a Minor Allele Frequency (MAF) >0.05 in the East Asian population. Candidate mRNAs were subsequently validated for their specificity in the target body fluids. We developed a targeted RNA sequencing assay incorporating these cSNPs, which was then used to detect and analyze case-type samples. **Results and conclusions:** Utilizing massive parallel sequencing (MPS), we developed a targeted RNA sequencing assay incorporating 31 cSNPs within 16 mRNAs. This assay was employed to identify the body fluid types in several case-type samples. Moreover, by comparing the cSNP genotypes of the samples with those of the suspect or victim, we were able to directly attribute each body fluid to the correct donor. Although the assay demonstrates significant promise, further validation tests are necessary. In conclusion, we have developed an MPS cSNP panel containing novel BFID cSNPs, and its detection capability for case-type samples has been validated.

Keywords
Forensic genetics; body fluid identification (BFID); Massively parallel sequencing (MPS); coding region SNP (cSNP); sexual assault.

1. Introduction

The identification of the origin of body fluids and tracing their donors are essential for reconstructing the sequence of events and determining the nature of criminal cases. In recent years, coding region single nucleotide polymorphisms (cSNPs) have demonstrated high specificity and the distinct advantage of directly linking certain body fluids to their respective donors in mixed samples [1–3]. In this study, we selected a set of body fluid identification (BFID) cSNPs to develop a targeted RNA sequencing assay. This assay was then applied to samples from a suspected rape case to evaluate its practical feasibility.

In this case, the victim reported being coerced into engaging in oral sex, which was subsequently followed by vaginal intercourse with the suspect, who allegedly employed a condom during the act of vaginal intercourse. The suspect, however, denied vaginal penetration. Due to the delay in reporting the incident to the police, crucial biological evidence from both the suspect and the victim was lost. Consequently, only a used condom and pieces of used tissues were submitted to us for testing.

2. Materials and methods

2.1 DNA and RNA isolation

After examining the condom, the inside and outside areas were swabbed first with a moist cotton swab, followed by a second, dry cotton swab. Both swabs were processed together to extract genomic DNA (gDNA) and total RNA using the QIAamp DNA Blood Mini Kit (QIAGEN, Hilden, Germany) and the mirVana™ miRNA Isolation Kit (Thermo Fisher Scientific, Waltham, USA). Sections of the tissue samples containing biological spots were excised to extract gDNA and total RNA as described above. Additionally, gDNA from the victim and suspect was extracted from their blood stains. Quantification of the DNA and RNA samples was performed using the Qubit™ dsDNA HS Assay Kit and the Qubit™ RNA HS Kit (Thermo Fisher Scientific), respectively. First-strand cDNA synthesis was carried out using the Transcriptor First Strand cDNA Synthesis Kit (Roche, Mannheim, Germany).

2.2 PSA test and STR genotyping

To detect prostate-specific antigen (PSA) on the inner and outer surfaces of the condom, as well as on the tissue samples, a rapid test strip (Huaxing Ruian Technology Co., Ltd, China) was employed. The short tandem repeats (STRs) in the aforementioned gDNA samples were genotyped using the VeriFiler PCR Amplification Kit (Thermo Fisher Scientific).

2.3 cSNP selection and targeted RNA/DNA sequencing

Body fluid-specific mRNAs containing cSNPs with a Minor Allele Frequency (MAF) >0.05 were screened in the East Asian population. These candidate mRNA-cSNPs were subsequently validated for their specificity in target body fluids. The Massive Parallel Sequencing (MPS) assay was then developed for gDNA and RNA, respectively. Pooled libraries were sequenced using the MiSeq Reagent Kit (Illumina, San Diego, USA) on the MiSeq platform (Illumina). All data were analyzed using a customized Python pipeline.

3. Results and discussion

The PSA test yielded negative results for both the condom and tissue samples. The STR genotypes obtained from both the inner (CI) and outer (CO) surfaces of the condom, as well as from the tissue samples, were found to be consistent with those of the victim.

In accordance with the protocols outlined in *Section 2.3*, a total of 31 cSNPs within 16 body fluid-specific mRNAs were identified, leading to the development of a targeted multiplex RNA MPS assay for BFID. The extracted RNA and DNA samples were sequenced using this newly developed RNA/DNA MPS assay, respectively. As illustrated in Figure 1, the saliva (64.93%) and semen (33.09%) biomarkers exhibited high read percentages (98.02%) for the CI sample, with minimal or no expression of other biomarkers. These results align closely with the victim's testimony regarding oral sex. For the CO sample, both the quality and quantity of the extracted RNA were suboptimal, resulting in poor RNA sequencing results of insufficient quality, which fell below the threshold required for reliable analysis. In the tissue samples, as depicted in Figure 2, the majority of reads (97.91%) were also observed in the saliva biomarkers.

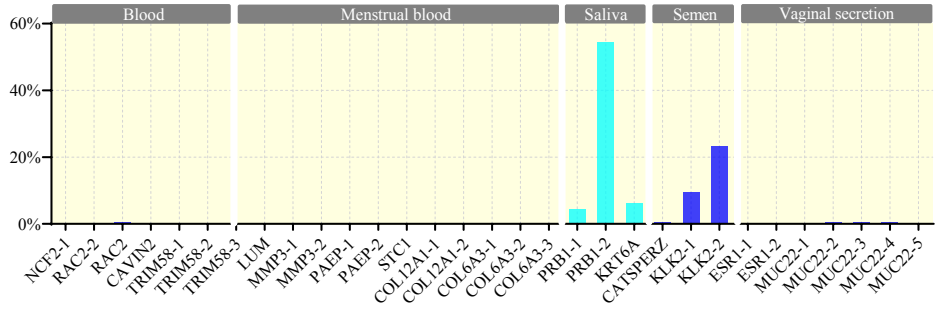


Figure 1. Percentage of reads in biomarkers of the inner surface of the condom (CI).

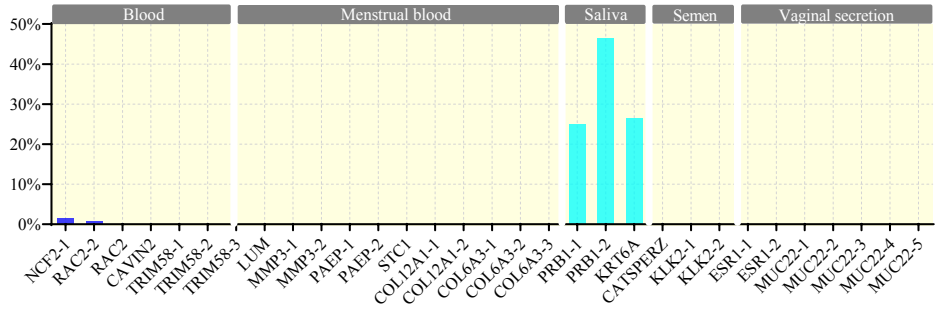


Figure 2. Percentage of reads in biomarkers of tissue samples.

In Table 1, the eligible cSNP genotypes of the CI sample, tissue, victim, and suspect are listed. These genotypes facilitate the assignment of each body fluid in the mixtures to the corresponding donor. The analysis reveals that the genotypes of three saliva biomarkers in both the CI sample and the tissue are consistent with those of the victim, while two biomarkers differ from those of the suspect. This finding supports the conclusion that the saliva originates from the victim and suggests the occurrence of oral sex. Additionally, there are two semen biomarkers in the CI sample that match those of the suspect, while two mismatches are observed between the CI sample and the victim. These results support the hypothesis that the suspect deposited the semen in the condom.

Table 1. Genotypic analysis of the cSNPs present in the samples

Specific Body fluid	Gene	cSNP	Victim	CI	Tissue	Suspect
Saliva	PRB1	rs10845341	***C/C***	***C/C***	***C/C***	C/T
	PRB1	rs76504934	***A/A***	***A/A***	***A/A***	A/C
	KRT6A	rs199613662	***C/C***	***C/C***	***C/C***	C/C
Semen	KLK2	rs198972	C/T	***C/C***	-	***C/C***
	KLK2	rs198977	C/T	***C/C***	-	C/C

Highlighted in bold and italic indicate a match of the observed genotype to the reference sample. Markers with reads below the threshold are shown with "-".

4. Conclusions

A customized MPS assay incorporating cSNPs for BFID was first applied to case-type samples. The findings revealed a mixture of semen and saliva on the interior surface of the condom, corroborating the victim's confession. However, before the cSNP data can be considered legally admissible in court, further validation studies and comprehensive investigations are required.

5. Acknowledgments

The authors would like to thank all the donors who donated body fluid samples. This study was supported by grants from the National Natural Science Foundation of China (82202080 and 82101971). The funding institutions had no role in study design, data analysis, publishing decisions, or manuscript preparation.

6. Conflict of interest statement

None.

References

[1] Hanson E, Dørum G, Zamborlin M, et al. Targeted S5 RNA sequencing assay for the identification and direct association of common body fluids with DNA donors in mixtures[J]. International Journal of Legal Medicine, 2023, 137(1): 13-32.

[2] Liu J, Cheng X, Liu F, et al. Identification of coding region SNPs from specific and sensitive mRNA biomarkers for the deconvolution of the semen donor in a body fluid mixture[J]. Forensic Science International: Genetics, 2021, 52: 102483.

[3] Johannessen H, Hanson E, Gill P, et al. Body Fluid Identification in Samples Collected after Intimate and Social Contact: A Comparison of Two mRNA Profiling Methods and the Additional Information Gained by cSNP Genotypes[J]. Genes, 2023, 14(3): 636.

[4] Sijen T, Harbison S. On the Identification of Body Fluids and Tissues: A Crucial Link in the Investigation and Solution of Crime[J]. Genes, 2021, 12(11): 1728.

[5] Dørum G, Bleka Ø, Gill P, et al. Source level interpretation of mixed biological stains using coding region SNPs[J]. Forensic Science International: Genetics, 2022, 59: 102685.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1165-1172
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Persistence of Proteomic Body Fluid Signals in DNA-workflows

Glendon Parker[1*], Allysa Tewksbury[2], Daniella Luistro[1], Kyle Burk[1]

[1]*Environmental Toxicology, University of California Davis,*
[2]*California Department of Justice - Riverside*

Abstract

Identification of the source of a biological stain adds important context to evidence and the corroboration of testimony. Current methods of body fluid identification however present significant challenges: they are presumptive, qualitative, and have sensitivity and specificity issues. When evidence is limited, investigators may have to choose between body fluid identification or maximizing DNA-isolation and human identification. These choices are based on incorrect assumptions about the persistence of protein signals in body fluid samples. A major assumption is that DNA isolation workflows, many of which rely on the Proteinase K (PK), fully eliminate protein as a source of forensic information. This study tested this hypothesis using the conditions of a DNA isolation workflow that were tested in the presence and absence of DTT and PK. In the absence of PK there was no change in protein quality or yields (n = 4, 101±21%). Discarded lysis buffers without PK in DNA-isolation workflows therefore are a rich source of proteomic information. As expected, PK digestion fully eliminated protein structures and Coomassie staining. However, the PK-digested peptides were detected at high levels. Semenogelin-1 peptides (634 ± 323 n = 3) that correspond to 87 ± 13 % of the protein were identified with a total signal of 7.6 ± 0.9 x 10^7 ions, that was 6 orders of magnitude greater than the lowest protein signal in the analysis level. We can therefore report that proteomic body fluid identification is fully compatible with current DNA isolation, *regardless of the protocol used*. Discarded material from DNA-workflows are a rich source of proteomic information.

Keywords

Forensic Proteomics, Body Fluid Identification, DNA-isolation, Proteinase K

1. Introduction

Case work investigation, particularly reconstruction of sexual assault, often needs more than just human identification, the individual source of DNA. Knowledge of the bodily source of DNA; the surface or orifice it originated from, is also important since this provides information about context, crime scene reconstruction, and confirmation of testimony. Obtaining this information faces intrinsic challenges however: sampling for DNA processing needs to be prioritized, sample size is limited and needs to be preserved for potentially subsequent analysis, and relevant biomolecules often degrade. When analyzing evidence of body fluid identification investigators have several options: they can use information encoded in nucleotide sources, such as DNA methylation patterns, or ribonucleic acid polymers, or they can analyze protein body fluid matrix through proteomics.

Body fluid protein patterns reflect the function and cellular origin of specific body fluids. These tissues are the result of multiple glands and cellular sources, that respond to the physiological cycles and biology of the body. Each cell type transcribes and translate around 11,000 genes with a median expression of about 600,000 copies per cell [1.]. In addition to cells that release contents into the extracellular space, cells in contact with a given body fluid may lyse and release internal contents. The result is that each body fluid sample contains a complex and dynamic biological record in its protein population.

Analyzing these populations has a deep history in forensic science. Prior to the DNA revolution, protein and epitope variants were a major resource for identifying source information [2.]. The legacy of these studies is an array of biomarker proteins that considered indicators of different body fluids [2, 3, 4.]. Antibodies to these proteins are used in immunochromatographic lateral flow devices (LFDs) to detect the presence or absence in casework [5.]. These devices are highly affordable, transportable and accessible. The fidelity of body fluid inference however is compromised. The most distinctive proteins are soluble and move around the body. Biomarker proteins may be expressed or at least detectable in multiple body fluids. Relying on one protein therefore results in false positives. Antibody specific may change with environmental conditions or degrade. These LFD tests are considered good presumptive evidence, but not for presentation in a legal setting. Reliable, validated, confirmatory-quality assays for body fluid are still in development. The lack of unambiguous confirmatory data is lacking, despite claims to the contrary [6, 7, 8.].

A major challenge to widespread body fluid analysis in casework is sample consumption. Investigators balance information gained against the irreversible

destruction of evidence. Human identification, DNA isolation and typing, takes precedence. Techniques that can analyze evidence after DNA has been isolated and purified therefore have an advantage. The three major alternatives to traditional serological approaches take different approaches. Epigenetic and transcriptomic approaches either use DNA remaining from casework [9.] or exploit mRNA nucleotides that pass-through DNA purification columns or magnetic beads [10, 11, 12.]. The other major technique to analyze body fluids, proteomic shotgun mass spectrometry, is under-utilized in forensic casework even though it has several major advantages: it is very sensitive, with a dynamic range of five to six orders of magnitude, and it is confirmatory, by relying on mass spectra that are precise, accurate, and easily validated [13.].

Part of the difficulty in implementing proteomic approaches for body fluid identification are common assumptions about nucleotide and protein-based sample processing [4.]. DNA isolation and proteomic shotgun mass spectrometry are widely thought to be incompatible. A pre-condition of DNA isolation protocols is the removal of protein, which in DNA protocols often requires Proteinase K (PK), an efficient and non-specific protease. Strong detergents and chaotropic agents that disrupt macromolecular structures and membranes are also used. PK results in almost randomly cleaved peptides that eliminate intact protein structures and are thought to be incompatible with proteomic analysis. Chemically, the detergents used are incompatible with liquid chromatography and mass spectrometry and must be removed from the sample.

Recent advances however challenge these assumptions. Protein sample processing methods are now available that can easily remove detergent and chaotropic agents. Improvements in computing and bioinformatics now make it easier to process complex peptide populations, even those digested with PK. Furthermore, advances in DNA-workflows, that use magnetic bead-based DNA-isolation such as Prep-Filer™ (Thermo Inc.) and DNA iQ™, have protocols that no longer use PK.

Experimental Design

To directly test the hypothesis that proteomics is not compatible with DNA-isolation, a study was conducted based on the magnetic bead DNA iQ™ DNA-isolation workflow that uses PK (and DTT) when semen is possibily in the sample. Samples of semen from three individuals were therefore incubated to replicate the DNA iQ™ pre-treatments in the presence and absence of reductant, PK and a 90-minute incubation at 56˚C. Coomassie staining from SDS-PAGE gels was used to evaluate the global stability of semen proteins. Proteomic shotgun mass

spectrometry was conducted to evaluate the proteomic content of PK-digested samples and observe if there was still proteomic information in the sample.

2. Material studied, methods, techniques

Frozen semen from a threes individual was sourced from Lee Biosolutions Inc. and as subjects using protocols approved by the University of California Davis (IRB# 2045384). DNA iQ™ reagents were purchased from Promega Corporation. Protein gels (SDS-PAGE) and reagents, including 12 well 4–20% Mini-PROTEAN® TGX™ Precast Protein Gels, and matching equipment were used (BioRad). Polyacrylamide gels were stained with colloidal G-250 protein stain [14.] and analyzed with densitometry with an Epson V500 Photo Scanner and ImageJ software (version 1.51s).

2.1. Sample Incubation

Semen protein was incubated (1 in 20 dilution of neat fluid) for 90 min at 56C in Differex™ Digestion Buffer in the presence and absence of 4.6 mM dithioerythritol and / or P K (17 µL) as described in the Differex™ System Technical Manual. At the t = 0 and t = 90 minute time point a 10 µL aliquot of the incubation was removed and applied to 10mL of 2X Laemmli Buffer then heated at 95°C in a water bath for 15 minutes. Another 10µL aliquot at each time point was also taken for application to mass spectrometry.

2.2.Mass Spectrometry

Samples were submitted to the University of California Davis Genome Center Proteomic Core Facility. Samples treated with PK were measured for peptide concentrations using a fluorescent peptide assay (Thermo Pierce) and 500 ng was applied to applied to a disposable Evotip C18 trap column (EvoSep BioSystems) as per manufacturer's instructions. The tipped samples were subjected to nanoLC on a Evosep One instrument (EvoSep BioSystems) using a preset set gradient at a rate of 60 samples per day. Peptides were directly eluted onto a hybrid trapped ion mobility spectrometry-quadrupole time of flight mass spectrometer (timsTOF HT, Bruker Daltonics) with a modified CaptiveSpray nano-electrospray ion source.

2.3. Data Analysis

Proteomic data was submitted to PEAKs software (version 11) with an ion-mobility module included. The data was processed using default settings with the human proteome used as a reference.

3. Results

Single semen samples from three individuals were treated to replicate the DNA iQ™ / Diffrex protocol for DNA-isolation, which included a 90-minute incubation at 56˚C in the presence of dithioerythritol (DTE) and PK. The treatment was modified to remove each, or both, of DTE, PK, or the 90-minute incubation (**Figure 1A**). Aliquots of each incubation as well as controls were applied to SDS-PAGE, stained in colloidal Coomassie stain, imaged (**Figure 1A**), and densitometry conducted (**Figure 1B**). From the SDS-PAGE gel the semen protein population is unaffected by the Diffrex™ buffers in the presence and absence of DTE and the 90-minute incubation. The incubation with all ingredients except PK had 101±21% densitometry of the original (n = 4). This endpoint of the incubation is the starting point of DNA-isolation for all samples except semen and hair [15.]. Protein therefore is fully preserved by the pre-treatment prior to DNA extraction.

These figures also demonstrate that PK incubation fully eliminated secondary and tertiary protein structures from the sample (**Figures 1A, 1B, lanes 3 & 4**) [16.]. Only the PK protein remaining to be detected through staining. This was expected and uncontroversial [16, 17.]. The consensus of the forensic, and much of the proteomic, field assumes that there is no longer any proteomic information in the sample. To test this, we removed the detergent from the sample and applied PK-digested peptides directly to proteomic shotgun mass spectrometry.

A proteomic shotgun mass spectrometry analysis resulted in 634 ± 323 peptides (**Figure 1C**, blue bars) that correspond to 87 ± 13 % of the semenogelin-1 sequence. The semenogelin signal was 7.6 ± 0.9 x 10^7 ions, 6 orders of magnitude greater than the lowest protein signal in the analysis. The semenogelin represented in the figure is from the lowest yielding replicate, with only 267 peptides. This demonstrates that PK digested samples still contain proteomic information that can be used for body fluid identification by orders of magnitude.

4. Discussion

Proteinase K does not eliminate proteomic information from evidence. This finding dramatically expands the scope of samples that can now provide body fluid identification through proteomics. PK treatments (Figures 1A, 1B, lanes 3 and 4) fully degrade secondary and tertiary structures but not peptides (Figure 1C). The dynamic range of 6 orders of magnitude in the sample demonstrates that even poorly preserved evidence samples should still be well over the lower limits of detection.

To test the hypothesis, this study used neat semen alone, from three individuals. A broader study of body fluids in DNA-workflows remains to be conducted. The study did not use mock SAE or cotton swabs, body fluid mixtures, or degraded samples. Three individuals are a small sample set and highly proteinaceous semen is ideal to test the hypothesis of information loss with PK, other fluids may differ. This study focused on steps in DNA-isolation that are prior to extraction of DNA. A systematic analysis of the proteome of DNA-isolation workflows, is being conducted.

5. Conclusion

This study demonstrates that Proteinase K digestion does not eliminate proteomic information and in fact the resulting peptides are a rich source of information including body fluid identification [18.]. The hypothesis of information loss is therefore invalidated. This finding would hold regardless of the workflow used and whether PK digestion is included or not in the protocol.

6. Acknowledgments

The authors acknowledge the UC Davis Forensic Science Graduate Program the research stipend. AT and DL thank thesis committee members Drs. Robert Rice and Ashley Hall for advice and reagents.

7. Conflict of interest statement

The authors have declared no conflict of interest.

8. References

[1] Meier F, Brunner AD, Koch S, Koch H, Lubeck M, Krause M, et al. Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. Molecular & cellular proteomics: MCP. 2018 Dec;17(12):2534-45.

[2] Sensabaugh G. Serology: Overview. In: Payne-James J, Byard RW, editors. Encyclopedia of Forensic and Legal Medicine (Second Edition). Oxford: Elsevier; 2016;254-66.

[3] Williams GA. Body fluid identification: A case for more research and innovation. Forensic Science International: Reports. 2020 2020/12/01/;2:100096.

[4] Parker GJ, McKiernan HE, Legg KM, Goecker ZC. Forensic Proteomics. Forensic Science International: Genetics. 2021.

[5] Bruijns B, Tiggelaar R, Knotter J, van Dam A. Use of Lateral Flow Assays in Forensics. Sensors (Basel). 2023 Jul 6;23(13).

[6] Leader MST. Molecular Serology Validation: Proteomic Body Fluid Assay Validation. In: Examiner NYCOotCM, editor. New York, NY; 2023;1-44.

[7] Legg KM. Validation of a Mass Spectrometry Based Serological Assay, Office USAR; 2016 12/1/2016. Report No.: 67484-LS-RIFContract No.: Document Number|.

[8] McKiernan HE, Brown CO, Arantes LC, Danielson PB, Legg KM. NextGen Serology: Leveraging Mass Spectrometry for Protein-Based Human Body Fluid Identification. Applications in Forensic Proteomics: Protein Identification and Profiling: ACS Publications; 2019;47-80.

[9] Harbison. S, R. F. Forensic body fluid identification: state of the art. Research and Reports in Forensic Medical Science. 2016;6:11-23.

[10] Procopio N, Bonicelli A. From flesh to bones: Multi-omics approaches in forensic science. Proteomics. 2024;n/a(n/a):2200335.

[11] Johannessen H, Hanson E, Gill P, Haas C, Bergseth EF, Ballantyne J, et al. Body Fluid Identification in Samples Collected after Intimate and Social Contact: A Comparison of Two mRNA Profiling Methods and the Additional Information Gained by cSNP Genotypes. Genes. 2023;14(3):636.

[12] Haas C, Neubauer J, Salzmann AP, Hanson E, Ballantyne J. Forensic transcriptome analysis using massively parallel sequencing. Forensic science international Genetics. 2021 May;52:102486.

[13] Goecker ZC, Legg KM, Salemi MR, Herren AW, Phinney BS, McKiernan HE, et al. Alternative LC-MS/MS Platforms and Data Acquisition Strategies for Proteomic Genotyping of Human Hair Shafts. bioRxiv. 2021:2021.03.15.435505.

[14] Neuhoff V, Arold N, Taube D, Ehrhardt W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. Electrophoresis. 1988;9(6):255-62.

[15] Promega Corporation. DNA IQTM System - Database Protocol. 2800 Woods Hollow Road, Madison, WI: Promega Corporation; 2016 Contract No.: Document Number|.

[16] Ebeling W, Hennrich N, Klockow M, Metz H, Orth Hd, Lang H. Proteinase K from Tritirachium album Limber. European Journal of Biochemistry. 1974;47(1): 91-7.

[17] Ahsan N, Mishra S, Jain MK, Surolia A, Gupta S. Curcumin Pyrazole and its derivative (N-(3-Nitrophenylpyrazole) Curcumin inhibit aggregation, disrupt fibrils and modulate toxicity of Wild type and Mutant α-Synuclein. Scientific reports. 2015 2015/05/18;5(1):9862.

[18] Van Steendam K, De Ceuleneer M, Dhaenens M, Van Hoofstat D, Deforce D. Mass spectrometry-based proteomics as a tool to identify biological matrices in forensic science. International journal of legal medicine. 2013 Mar;127(2):287-98.

Tables and Figures (to be embedded in text AND uploaded as separate source files)

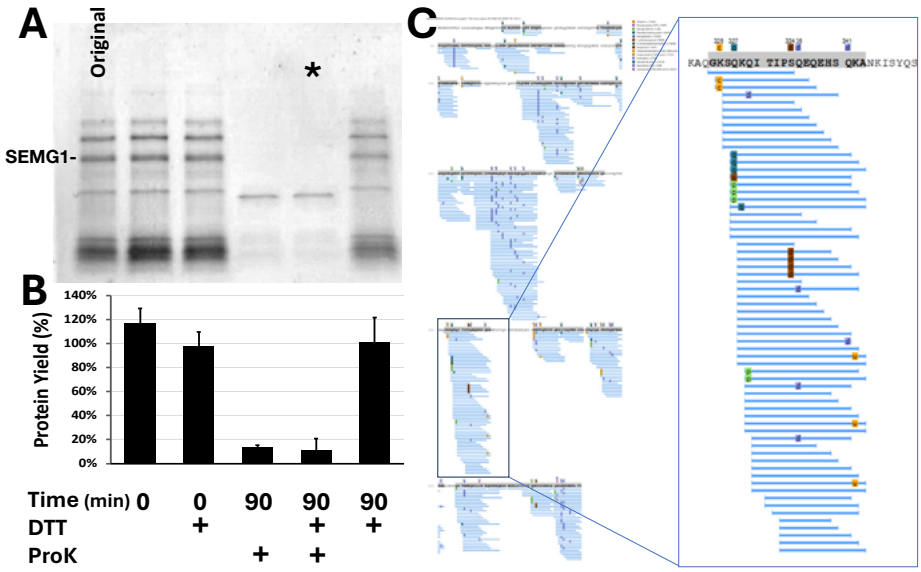


Figure 1. Proteinase K Degrades But Does Not Eliminate Protein.
Semen protein was incubated for 90 min in the DNAiQ™ Incubation buffer with ± DTT and ± ProK following the Differex™ System Technical Manual (n = 3). A) Resulting samples were applied to SDS-PAGE gels and (B) densitometry conducted. A ProK treated sample (*) was applied to the EvoSep timsToF Mass Spectrometer. Data was analyzed using PEAKs software. C) Detected peptides (blue bands) from semenogelin 1 protein (SEMG1) in the least abundant sample are shown for both the whole protein and an insert to illustrate protein coverage.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1173-1179
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Reverse Body Fluid Identification Workflow: A Direct to DNA Approach

Maria-João Porto[1*], Joana Ferreira[2], Ana Bento[1], Vanessa Bogas[1], Virginia Lopes[1], Lisa Sampaio[1], Nair Gouveia[1], Francisco Corte-Real[1,3], António Amorim[1,3,5]

[1]*Instituto Nacional de Medicina Legal e Ciências Forenses, I.P., Portugal.*
[2]*Master student of Faculdade de Medicina da Universidade de Coimbra, Portugal.*
[3]*Faculdade de Medicina da Universidade de Coimbra, Portugal.*
[4]*Faculdade de Ciências da Universidade de Lisboa, Portugal.*
[5]*REQUIMTE, Analytical Development Group, Laboratório Associado FCT, Portugal.*

** Corresponding author at: Instituto Nacional de Medicina Legal e Ciências Forenses, I.P., Portugal, m.joao.porto@inmlcf.mj.pt*

Abstract

Preliminary serological assays to detect the presence of semen in sexual assault offenses are often used by forensic laboratories as a screening method prior to DNA analysis, in order to select the best samples for extraction and determine if the victim's collected items are suitable for differential extraction. However, in many situations, samples have low amounts of semen or other male bodily fluids which can generate a semen negative result and, for that reason, some samples may not be selected for DNA profiling. To avoid this lack of information, in our laboratory until the end of 2022 and whenever possible, samples were processed in separate workflows, one to determine the presence of biological fluids and the other to generate the DNA profile. Due to the high percentage of samples detected with negative preliminary tests and without eligible DNA profiles for report purposes, we conducted a study to better quantify and understand those results. A total of 657 samples from sexual assault crimes were then analysed and it was detected that 60.4% of them had, for different reasons, no information relevant for the case under study. Given these results, a more efficient workflow (time and cost effective) was proposed to be applied to all forensic samples, reaching first a reportable DNA profile by applying a *Direct to DNA* approach followed by serological assays for body fluid identification, if necessary.

Keywords

Sexual assault, Body fluid identification, Semen detection, *Direct to DNA* approach.

1. Introduction

When forensic DNA laboratories receive evidence from a crime scene their first task is to check for the presence of biological material, namely blood, semen and/or saliva; the same principle is applied to items related to victims from sexual assault cases, which may include internal or external genitalia swabs, oral swabs, skin swabs, fingernail swabs, condoms, underwear and other clothing, towels or bedding collected in the crime scene, etc. Examination of the exhibits by naked eye or using a forensic light source is done in order to detect the presence of body fluid stains and many laboratories perform preliminary assays on items where biological material is potentially present before sending a cutting or swab for extraction and subsequent DNA typing. Nevertheless, this methodology may fail the detection of DNA profiles useful for solving the criminal case, since the recent DNA typing kits are more sensitive than many serology tests used by laboratories. Other laboratories perform a differential extraction without prior screening for the presence of semen, namely in internal swabs collected during the examination of victims of alleged sexual assaults, but male DNA can be lost during the separation between both epithelial and sperm fractions; on the other hand, some samples do not contain semen but other biological fluids from the perpetrator, which leads to a larger number of samples that need DNA profiling [1, 2, 3, 4, 5, 6].

To identify the presence of bodily fluids, our laboratory has implemented presumptive and/or confirmatory assays to detect semen, blood and saliva and, until the end of 2022, all samples selected for DNA extraction and posterior quantification/amplification were also tested to determine the type of biological evidence in question (when enough sample was available) in different workflows.

For semen identification, all presumptive positive results were then tested in order to visualize sperm cells. However, in sexual assault cases there are many samples with a semen presumptive positive result but with a negative confirmatory test, meaning that this biological fluid cannot be confirmed. It was also detected that in several situations the analysed samples did not present probative DNA results and, consequently, it would not have been necessary to test them for the presence of bodily fluids.

The aim of this study was to evaluate the methodology implemented in sexual assault samples regarding the need to perform semen preliminary tests in all samples, and propose a more efficient workflow to be applied to forensic samples received in our laboratory.

2. Materials and methods

A total of 647 samples from sexual assault crimes occurred between 2020 and 2021 were selected from female and male victims (cases already closed); the majority of the analysed samples included swabs, clothing and bed sheeting. A Crime-lite® 82S (foster + freeman) with blue light and wavelength 420-470nm was used as an alternative light source to detect suspicious stains on clothes. Whenever possible, samples that needed semen identification (stains on fabric or swabs) were subsampled by making three cuts: one to detect the presence of semen, one for non-differential extraction and one for differential extraction, if necessary. All samples were anonymized in order to not identify their donors or even the case.

Two parallel and independent workflows were applied to all selected samples: one to detect the presence of semen and the other to obtain the male genetic DNA profile of interest.

2.1. Semen detection

As a presumptive test for semen detection it was used the Seratec® Prostate-specific Antigen (PSA) Semiquant immunochromatographic rapid test (Göttingen), also used as a screening method since only positive PSA results were sent for microscopic visualization of spermatozoa using the Christmas Tree (CT) staining method [7].

2.2. DNA profiling

DNA extraction was performed with the PrepFiler™ Forensic DNA Extraction Kit (Applied Biosystems), with an elution volume of 50µL, and then quantified using the Quantifiler® Trio DNA Quantification kit (Applied Biosystems). The screening of samples that required differential extraction was done by evaluating the quantification results: in general, if the male-female DNA ratio is greater than 1:10 and the male component is above 0.5ng/µL (indicative value, which may be lower), a differential extraction was made using the Sampletype i-sep® DL-MB columns (Biotype) with subsequent purification with the PrepFiler kit followed by another quantification. Only the sperm fraction of the differential extraction was amplified.

Quantification results from both differential and non-differential extraction were then grouped into three categories according to the existing amount of male DNA: greater than or equal to 0.1 ng/µL (where DNA profiles are generally easily achieved); between 0.01 ng/µL and 0.1 ng/µL (where DNA profiles may or cannot

be obtained); lower than 0.01 ng/µL (where, in many simples, it is not possible to obtain a valuable DNA profile or any profile at all). When multiple cuts with identical importance were made from the same evidence (mainly items of clothes), only the best ones were selected for amplification (male-female ratios were taken into account and samples without male DNA were not amplified).

Eligible samples were then amplified with the GlobalFiler™ PCR Amplification Kit (Applied Biosystems) in order to obtain an useful autosomal STR profile for solving the case. When very low levels of male DNA were present (below 0.01 ng/µL) or whenever the male-female ratio did not allowed an autosomal STR profile (in most situations with ratios above 1:10), the Yfiler® Plus PCR Amplification Kit (Applied Biosystems) was used to detect a Y-STR haplotype.

All kits were performed according to the manufacturer's protocol.

3. Results

The results revealed that 391 out of 647 samples (60.4%) had no relevance for the cases analysed. This percentage includes samples without male DNA, with DNA profiles or haplotypes that could not be evaluated (namely with the majority of the alleles below stochastic level) or with DNA profiles from male victims. Redundant samples (collected, for example, from the same item of clothing) are also included in this group; only samples with the highest concentrations of male DNA and/or the best male-female ratios were selected for DNA profiling.

The remaining 256 samples (39,6%) had a reportable DNA profile or haplotype that could be considered probative for the criminal case, with 90.7% of them (232 samples) with male DNA greater than or equal to 0.01 ng/µL (Fig. 1). The majority of samples presented a complete genetic autosomal genetic profile or Y-STR haplotype, with a higher percentage of incomplete profiles in samples with male DNA below 0.01 ng/µL, as expected.

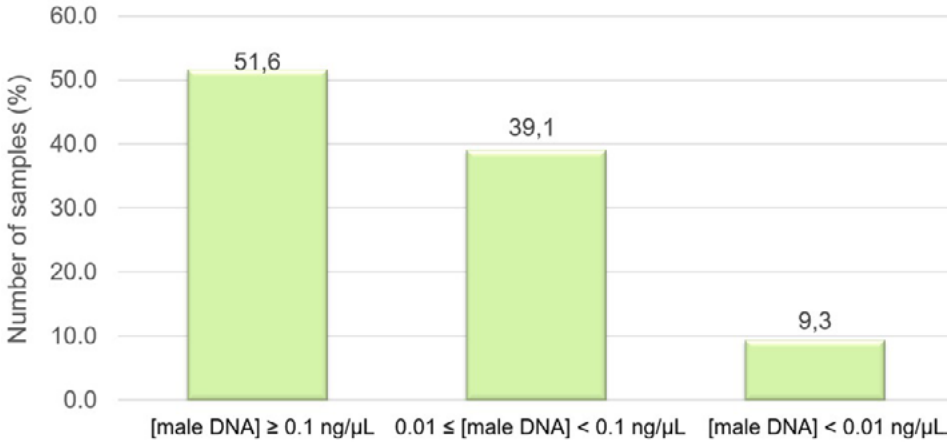
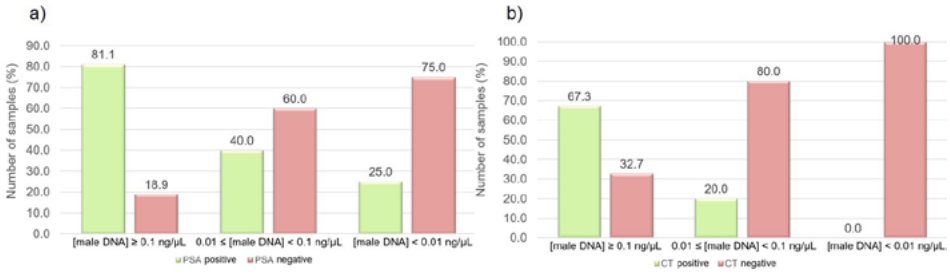


Figure 1. Number of samples (%) with reportable male DNA profiles/haplotypes distributed by male DNA concentration.

For samples with reportable DNA profiles, only 153 tested positive for PSA (107 of them with the higher concentration of male DNA) and a positive CT result was achieved in approximately half of them (80 samples), all with male DNA greater than or equal to 0.01 ng/µL. The results (in percentage) distributed according to the male DNA concentration are shown in Fig. 2.



Figre 2. Number of samples (%) with reportable male DNA profiles/haplotypes, distributed by male DNA concentration: a) PSA results and b) CT results.

It was also observed that in the group with samples without relevance for the cases under study, 83 samples (21,2%) showed a positive result for PSA even with reduced concentration (less than 0.01 ng/µL) or absent male DNA.

4. Discussion

The quantification step is essential for the selection of forensic samples for amplification. Only samples with male DNA and with the best male-female DNA ratios are selected to perform a genetic DNA profile and, as expected, higher DNA concentrations revealed a higher percentage of complete DNA profiles and positive PSA and CT results. However, in the majority of the analysed samples (60.4%), the perpetrator's male DNA was not present or could not be assessed (especially in samples with a lower concentration of male DNA). From our point of view, in these situations, carrying out presumptive or confirmatory tests to detect semen is unnecessary since the absence of a genetic DNA profile does not allow the aggressor to be identified.

It was also observed that, in many situations, a male DNA genetic profile was not associated to a positive PSA result. These results can be explained by the presence of the aggressor´s saliva or epithelial cells, which may be present in sexual assault crimes, and also because in some cases the victim is male and the genetic profile obtained is from his own DNA (unrelated to semen).

Negative CT results (for semen confirmation) are observed in almost half of positive PSA results and may be related to: lower concentration of male DNA observed in several samples, degradation of spermatozoa that have become unsuitable for microscopic visualization, or because samples originate from azoospermic or vasectomized individuals. Another explanation may be the presence of PSA in some female body fluids which can generate a positive result and, consequently a negative CT result [2, 4, 8]. On the other hand, in samples with very low concentrations of male DNA it is still possible to obtain a male reportable genetic profile/haplotype (even if incomplete) and this may be due to the high sensitivity of the current amplification kits.

5. Conclusion

If a *Direct to DNA* approach was adopted on the studied samples (where DNA analysis is performed prior to identification of bodily fluids) only 39.6% of them would have been tested for the presence of semen, saving time, costs and personnel efforts, since microscopic observation is laborious and time-consuming. This workflow (currently implemented in our laboratory) is more efficient, allows standardization of the implemented techniques and, above all, no loss of information relevant to the judicial process was detected.

6. Conflict of interest statement

The presented results are included in the master dissertation of the co-author JF.

7. References

[1] J.M. Butler, Advanced Topics in Forensic DNA Typing; Methodology, Elsevier Academic Press, San Diego, 2011.

[2] Forensic DNA Applications. An Interdisciplinary perspective. Edited by Dragan Primorac and Moses Schanfield. CRC Press, 2014.

[3] SWGDAM Guidelines for the Collection and Serological Examination of Biological Evidence, 2015, https://www.swgdam.org/_files/ugd/4344b0_b3deba7a272b4b268d7f522840607410.pdf.

[4] K. Sakurada, K. Watanabe, T. Akutsu, Current Methods for Body Fluid Identification Related to Sexual Crime: Focusing on Saliva, Semen, and Vaginal Fluid. Diagnostics. 10 (2020) 693, https://doi:10.3390/diagnostics10090693.

[5] G. Alderson, H. Gurevitch et al., T. Casimiro, Inferring the presence of spermatozoa in forensic samples based on male DNA fractionation following differential extraction, Forensic Sci. Int. Gnet. 36 (2018) 225-232, https://doi:10.1016/j.fsigen.2018.06.014.

[6] SWGDAM Recommendations for the Efficient DNA Processing of Sexual Assault Evidence Kits, 2016, https://www.swgdam.org/_files/ugd/4344b0_4daf2bb5512b4e2582f895c4a133a0ed.pdf.

[7] J.P. Allery, N. Telmon, R. Mieusset, et al., Cytological detection of spermatozoa: comparison of three staining methods, J. Forensic Sci. 46 (2001) 349-351. https://doi:10.1520/jfs14970j.

[8] K. Unuma, H. Sato, S. Wen, Y. Makino, A. Hirakawa, K. Uemura, The proportion of false-positives in positive Seratec® prostate-specific antigen SemiQuant test results in postmortem screening for seminal fluid. Legal Medicine 62 (2023) 102243, https://doi:10.1016/j.legalmed.2023.102243.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1181-1187
DOI: https://dx.doi.org/10.15304/cc.2025.1869

PSA and Semenogelin effectiveness to detect semen in sexual assault investigations

Mercedes Aler[1], María J. Castagnola[2], Francisco Medina-Paz[2], Maik Schlieper[3], Christian Stadler[3], Maria Isabella Negretti[3], Nora Ronan-Milz[4], Polina Khalizov[2], Gabriela Roca[3], Sara C. Zapico[2,5*]

[1]*Universidad de Valencia. Fornoles 3, 46015, Valencia, Spain*

[2]*New Jersey Institute of Technology, Department of Chemistry and Environmental Sciences, 161 Warren St., Tiernan Hall 365, Newark, NJ 07102, USA.*

[3]*SERATEC® Gesellschaft für Biotechnologie mbH, Ernst-Ruhstrat-Strasse 5, 37079 Göttingen, Germany.*

[4]*Lehigh University, 27 Memorial Drive West, Bethlehem, PA 18015 USA.*

[5]*Smithsonian Institution, National Museum of Natural History, Anthropology Department and Laboratories of Analytical Biology, 10th and Constitution Ave, NW, PO Box 37012, Washington DC 20560, USA.*

** Corresponding author at: New Jersey Institute of Technology, Department of Chemistry and Environmental Sciences, 161 Warren St., Tiernan Hall 365, Newark, NJ 07102, USA and Smithsonian Institution, National Museum of Natural History, Anthropology Department and Laboratories of Analytical Biology, 10th and Constitution Ave, NW, PO Box 37012, Washington DC 20560, USA. E-mail: sc338@njit.edu* sc338@njit.edu *and casado-zapicos@si.edu*

Abstract

Semen detection at crime scenes is essential for sexual assault investigations. There are several approaches to detecting body fluids at crime scenes. Lateral flow immunochromatographic (LFI) tests, based on antigen-antibody reactions, are widely used for this purpose. Currently, there are two types of LFI for the detection of semen: those based on prostate-specific antibodies (PSA), produced in the prostatic epithelium; and those based on semenogelin (Sg), produced in the seminal vesicles, the vas deferens, the prostate and the epididymis, both of which are secreted in semen. The aim of this study was to assess the effectiveness of PSA and Sg LFI tests to detect seminal fluid, evaluating the factors that could affect the identification of these proteins and assessing the possibility of obtaining DNA. To do this, PSA and Sg LFI tests were analyzed in four types of semen samples: single and pool donors, vasectomized and 30-year-old frozen pool

donor (to simulate an unresolved case). Serial dilutions of the samples were carried out from 1:2 to 1:32000. Our results indicated that both tests were positive in all samples, except that Sg began to be negative from 1:2000 in the 30-year-old group sample and from 1:8000 in the rest of the samples. Conversely, PSA began to be negative from 1:32000 in the fresh pool donor and from 1:32000 in the 30-year-old group, indicating better PSA performance than Sg, even in a vasectomized sample. These results were independent of the presence of sperm cells, still being positive without them. Regarding DNA analysis from 1:2000 onwards, it was not possible to obtain quantifiable DNA to perform the STR analysis. PSA was only detected in other dilutions of male urine samples. Future research will aim to correlate the positivity of these tests with DNA concentration and identifying the Y chromosome for the presence/absence of male, to arrive at an unambiguous and powerful identification.

Keywords
Semen, PSA, Semenogelin, urine, DNA

1. Introduction

Body fluid identification represents one of the key components in criminal investigation to find biological evidence related to the crime, potentially identify the victim and/or perpetrator, and reconstruct the crime events[1].

In sexual assault cases, the identification of semen is crucial to prove the assault and identify the perpetrator. In this regard, among the body fluid identification assays to perform, lateral flow immunochromatographic (LFI) tests have been widely used at the crime scene and in the lab due to their specificity and speed of obtaining results[2]. The tests use a mobile and stationary monoclonal antibody against the protein of interest, which will form a visible pink line if the protein is present[3]. In semen, some of these tests are based on testing prostate-specific antigen (PSA), a protein produced in considerable amounts only in the prostate, being able to detect in semen from azoospermic males[4]. Other LFI tests use semenogelin, a protein that originates in the seminal vesicles, and a substrate for the PSA to identify seminal plasma in forensic samples[5].

Based on these premises, the aim of the present work was to assess the effectiveness of PSA and semenogelin to detect semen, also evaluating the factors that could impact this detection, and trying to obtain DNA.

2. Material studied, methods, techniques

2.1. Samples

Four types of semen samples were assessed: 1) Single human donor; 2) Pooled human donors; 3) Vasectomized; 4) 30-year-old sample from a single donor. Dilutions of these samples were carried out from 1:2 to 1:32000 in water and extraction buffer provided by the LFI tests. Other body fluids were tested to evaluate specificity: breast milk, urine, saliva, sweat, and vaginal fluid. The New Jersey Institute of Technology Institutional Review Board (IRB) approved the protocol related to human body fluid experimentation (Protocol number 2110013076).

2.2. Methods

— *Lateral Flow Immunochromatographic tests:* SERATEC® PSA Semiquant (SERATEC®, Göttingen, Germany); RSID™ Semen, detects Sg (RSID™, Lombard, IL, USA), following the manufacturers' instructions.
— *Cytology: To visualize and count the sperm cells, Christmas Tree stain and direct counting without stain were carried out.*
— *DNA extraction:* DNA extraction was performed with Prepfiler DNA Extraction Kit (Thermo Fisher Scientific, Waltham, MA, USA).
— *DNA quantification:* DNA concentration was evaluated with human-specific quantitation, Quantifiler Trio in QuantStudio instrument (Thermo Fisher Scientific, Waltham, MA, USA).

3. Results

3.1. Comparative identification of the two proteins in four seminal fluids.

Our results indicated that both tests were positive in all samples, except that Sg began to be negative from 1:2000 in the 30-year-old group sample and from 1:8000 in the rest of the samples. Conversely, PSA began to be negative from from 1:32000.

3.2. Positiveness of PSA is independent of the number of sperm cells.

To assess if the number of sperm cells was related to the positiveness of the tests, samples and their dilutions were stained with Christmas Tree to count the number of cells. According to our results, these two factors are independent. From the dilution 1:1000 and up, where no sperm cells were found, the tests were still strongly positive.

3.3. Possibility of DNA profiling

DNA concentration and quality was evaluated to determine the possibility of obtaining a DNA profile from the samples and their dilutions. DNA concentration was enough to develop STR and Y-STR profiles, and the quality of the DNA acceptable up to 1:1000-1:2000 dilution. Beyond this dilution, both concentration and quality of the DNA drops, probably due to the lack of sperm cells, as described previously, though PSA remains positive.

3.4. PSA assessment in other body fluids.

All body fluids evaluated were negative for PSA, except two vaginal fluid samples, which became negative after dilution. Additionally, three male urine samples were still lightly positive at 1:1000 dilution, becoming negative at 1:1050 dilution (Table 1).

Table 1. Evaluation of false positives of PSA in other body fluids.

Body Fluids	Positive when pure or diluted	Extra-dilutions with negative results
Breast Milk Single and Pool	Never	From pure to 1:32000
13 Urine Female Single	Never	From pure to 1:32000
2 Nose Secretion Single	Never	Pure
Sweat Single 01	Never	1:10-1:500
Sweat Single 02	Never	Pure-1:32000
Saliva	Never	1:10-1:100
Vaginal Fluid Single 01	Never	1:100
Vaginal Fluid Single 02	Pure, 1:2, 1:10	1:100, 1:1000, 1:32000
Vaginal Fluid Pool 03	Pure, 1:2, 1:10	1:100, 1:1000, 1:32000
Urine Male Liquid Single	Pure, 1:100	1:300, 1:1000
7 Urine Male Single	Pure, 1:16, 1:100	1:256
3 Urine Male Single	Pure, 1:1000	1:1050

4. Discussion

Results of the present study indicated that both tests were effective to detect semen, though Semenogelin was less sensitive in higher dilutions. These findings are in agreement with the study of Boward and Wilson[6], in which they compared PSA detection through ABAcard® tests with RSID™-Sg, demonstrated the lower sensitivity of the later, but also, in post-coital samples, indicating that PSA is more

sensitive than Sg, proven more effective in vasectomized samples. DNA concentration was enough to develop STR and Y-STR profiles, and the quality of the DNA acceptable up to 1:1000-1:2000 dilution. Additionally, this study assessed the detection of both proteins in other body fluids, reaching the same conclusions as our present work (Figure 1). The study of Stroud et al.[7] also evaluated positive PSA results in other fluids, confirming our findings that some male urine could retrieve positive results. When urine is pure, it is expected to encounter traces of seminal fluid. It is also possible to find PSA in pure vaginal fluid, which could mislead the correct identification of seminal fluid, however, dilution of the sample will reduce the possibility of a false positive.

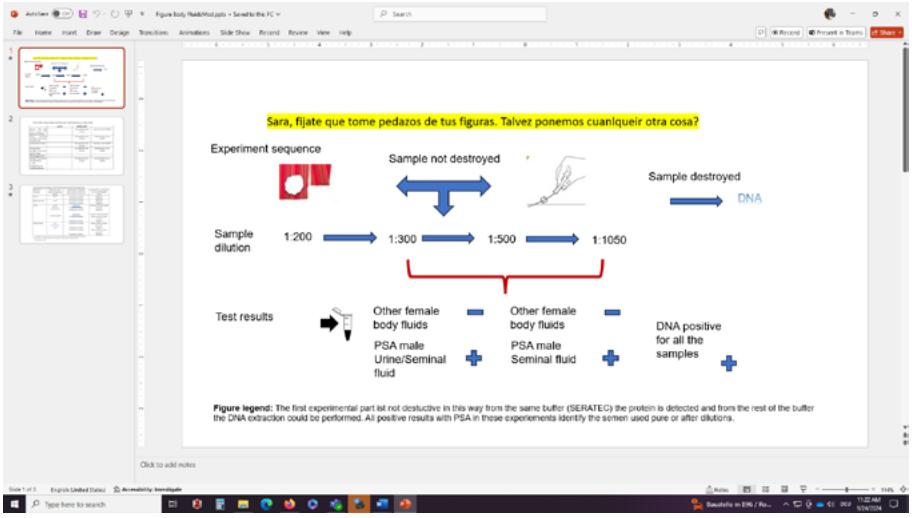


Figure 1. Summary of the results.

5. Conclusion

This is the first extensive study assessing the effectiveness of PSA and Sg to detect semen in different types of samples. According to our results, PSA seems more sensitive than Sg in higher dilutions of semen, being independent of the presence of sperm cells. Additionally, it is possible to obtain DNA for STR analysis up to 1:2000 dilution. Also, our findings indicate that no false positives were encountered for PSA, except some male urine samples, which required their dilution. A summary

of the results is depicted in Figure 1. Future research aims to expand these results, trying to correlate the positiveness of these tests with the retrieval of full STRs and Y-STRs profiles, as well as with mixture of other body fluids to simulate a sexual assault scenario.

6. Acknowledgments

The authors acknowledge Alexander Griberman for his revision of the manuscript and valuable comments. The authors would like to thank SERATEC® (SERATEC®, Göttingen, Germany), for providing the kits to carry out this work. María Josefina Castagnola is funded by the Fulbright Foreign Student Program from the Ministry of Education of Argentina.

7. Conflict of interest statement

Dr. Christian Stadler, Dr. Gabriela Roca and Maik Schlieper are employees of SERATEC®.

8. References

[1] Sijen T, Harbison S: On the Identification of Body Fluids and Tissues: A Crucial Link in the Investigation and Solution of Crime. *Genes (Basel)* 2021, **12**(11).

[2] Zapico SC, Roca G: Making the Most of Lateral Flow Immunochromatographic Tests: An Efficient Protocol to Recover DNA. *Methods Protoc* 2024, **7**(1).

[3] Sakurada K, Watanabe K, Akutsu T: Current Methods for Body Fluid Identification Related to Sexual Crime: Focusing on Saliva, Semen, and Vaginal Fluid. *Diagnostics (Basel)* 2020, **10**(9).

[4] Khaldi N, Miras A, Botti K, Benali L, Gromb S: Evaluation of three rapid detection methods for the forensic identification of seminal fluid in rape cases. *J Forensic Sci* 2004, **49**(4):749-753.

[5] Sato I, Kojima K, Yamasaki T, Yoshida K, Yoshiike M, Takano S, Mukai T, Iwamoto T: Rapid detection of semenogelin by one-step immunochromatographic assay for semen identification. *J Immunol Methods* 2004, **287**(1-2):137-145.

[6] Boward ES, Wilson SL: A comparison of ABAcard((R)) p30 and RSID-Semen test kits for forensic semen identification. *J Forensic Leg Med* 2013, **20**(8):1126-1130.

[1] Stroud A, Gamblin A, Birchall P, Harbison S, Opperman S: A comprehensive study into false positive rates for 'other' biological samples using common presumptive testing methods. *Sci Justice* 2023, **63**(3):414-420.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1189-1194
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Identification of Postmortem Blood through detection of D-dimer

Amy N. Brodeur[1], Tai-Hua Tsai[1], Huxi Wang[1], Gulnaz T. Javan[2], Gabriela Roca[3], Sara C. Zapico[4,5*]

[1]*Boston University Chobanian & Avedisian School of Medicine, 72 East Concord Street, R806, Boston, MA 02118, USA.*

[2]*Alabama State University, Department of Physical and Forensic Sciences, Montgomery, AL 36104, USA*

[3] *SERATEC® Gesellschaft für Biotechnologie mbH, Ernst-Ruhstrat-Strasse 5, 37079 Göttingen, Germany.*

[4]*New Jersey Institute of Technology, Department of Chemistry and Environmental Science, Newark, NJ, 07102, sc338@njit.edu.*

[5]*Smithsonian Institution, National Museum of Natural History, Anthropology Department and Laboratories of Analytical Biology, Washington, DC, 20560, casado-zapicos@si.edu.*

** Corresponding author at: New Jersey Institute of Technology, Department of Chemistry and Environmental Science, Newark, NJ, 07102, sc338@njit.edu; and Smithsonian Institution, National Museum of Natural History, Anthropology Department and Laboratories of Analytical Biology, Washington, DC, 20560, casado-zapicos@si.edu*

Abstract

There are different approaches to identifying blood at a crime scene. Depending on the scenario, it may be important to determine whether a person was alive at the time of blood deposition. We focus on the fibrinolysis pathway to identify potential biomarkers for postmortem blood. Fibrinolysis is the natural process that breaks down blood clots after healing a vascular injury. Among the fragments produced in fibrinolysis are soluble fibrin monomers crosslinked between two adjacent outer D domains detectable in plasma, known as D-dimers. Currently, D-dimer detection is used to distinguish menstrual blood from peripheral blood, as fibrinolysis plays an essential role in menstruation. Based on the rapid onset of fibrinolysis after death, D-dimer could be considered a biomarker for postmortem blood. SERATEC® has developed the PMB immunochromatographic assay to detect human hemoglobin and D-dimer simultaneously. This study aimed to assess the possibility of using the PMB test to detect postmortem blood. In addition, assessments of D-dimer levels in peripheral, menstrual, and postmortem blood were carried out, and the

ability to obtain STR profiles from postmortem blood was evaluated. Our results showed that all postmortem samples reacted positively for the presence of hemoglobin using the SERATEC® PMB tests, and all but one degraded sample showed a positive result for D-dimer. All peripheral blood samples from living individuals showed positive results for hemoglobin and negative results for D-dimer detection, except for one sample with a weak positive. Menstrual blood samples gave positive results for hemoglobin but variable results for D-dimer. The DIMERTEST® Latex assay was used for semi-quantitative measurement of D-dimer concentrations, with postmortem and menstrual blood yielding higher D-dimer concentrations compared to antemortem peripheral blood. Full STR profiles were developed for all postmortem samples tested except for one degraded sample, pointing to not only detecting postmortem blood at the crime scene but also the identification of the victim.

Keywords
Postmortem Blood, Menstrual Blood, D-Dimer.

1. Introduction

Blood is one of the most encountered biological samples in a criminal investigation. Its correct detection could lead to the identification of the perpetrator and/or victim [1]. Depending on the forensic case, apart from identifying blood, it could also be important to test for postmortem indicators in blood to determine whether the person was alive at the time of the blood deposition when a pool of blood is left at the scene.

A potential approach to identifying postmortem blood relies on the fibrinolysis pathway [2]. Fibrinolysis is the natural process that breaks down blood clots after healing a vascular injury. Among the degradation products of this pathway, the soluble fibrin monomers crosslinked between two adjacent outer D domains, known as D-dimers, are detectable in human plasma [3].

Currently, the detection of D-dimer is utilized to distinguish menstrual blood from peripheral blood in the forensic field. Fibrinolysis plays an important role in menstruation [4]. This feature has been used to distinguish menstrual blood from peripheral blood in cases of alleged sexual assault through an immunochromatographic multiplex assay [5].

Apart from menstrual blood, D-dimer is also elevated in postmortem blood. SERATEC® (SERATEC®, Göttingen, Germany) has developed an immunochromatographic assay to simultaneously detect human hemoglobin and D-dimer, called PMB. This test has been primarily developed to presumptively identify menstrual blood[5].

Based on the aforementioned, this study aimed to assess the possibility of using the PMB test to detect postmortem blood and generating STR profiles from these samples.

2. Material studied, methods, techniques

2.1. Samples

Postmortem Blood Samples (PM): forty samples derived from blood and 40 human corpses from criminal cases (homicide, suicide, and overdose) with postmortem intervals between 4–91 hours were collected.

Antemortem Blood Samples (AM): 21 AM liquid samples.

Menstrual Blood Samples (MB): 10 MB samples.

2.2. Methods

— Lateral Flow Immunochromatographic Tests: SERATEC® PMB test (SERATEC® GmbH, Göttingen, Germany) simultaneously detects human hemoglobin (limit 20 ng/ml) and D-dimer (limit of 400 ng/ml).
— Rapid Latex Agglutination Assay: DIMERTEST® Latex assay (Siemens Healthineers, Erlangen, Germany).
— Total DNA Extraction: DNeasy Blood and Tissue Kit (Qiagen®) protocol (Qiagen, Germantown, MD, USA).

DNA Quantification: PowerQuant System (Promega Corporation, Madison, WI) in QuantStudio (Thermo Fisher Scientific, Waltham, MA, USA).

STR Profiling: Promega PowerPlex® Fusion 6C System (Promega Corporation, Madison, WI) in SeqStudio ((Thermo Fisher Scientific, Waltham, MA, USA).

3. Results

3.1. D-dimer detection by the PMB test

D-dimer evaluation with the SERATEC® PMB test (SERATEC® GmbH, Göttingen, Germany) demonstrated that all the PM samples except one (degraded, foul-smelling) were positive for D-dimer. Menstrual blood samples gave variable results for D-dimer, and only one AM blood sample gave a positive result for D-dimer. Please see the summary of these results in Table 1.

3.2. D-dimer quantification by Rapid Latex Agglutination Assay

According to the results of this test, PM and MB samples presented concentrations of D-dimer between 12-100 mg/L, while AM samples D-dimer concentrations were below 1.6 mg/L (Table 1).

3.3. STR profiling from PM blood

It was possible to obtain full DNA profiles from all PM blood except one degraded sample (Table 1).

Table 1. Summary of the results. PM, postmortem blood; MB, menstrual blood; AM, antemortem blood samples; N.A., not applicable.

Sample	PMB test (number of positive D-dimer samples/total samples)	Rapid Latex Agglutination Assay for D-dimer (concentrations in mg/L)	STR profile
PM	39/40	12-100 mg/L	39/40
MB	5/10	12-100 mg/L	N.A.
AM	1/21	Below 1.6 mg/L	N.A.

4. Discussion

All postmortem samples except one showed a positive result for D-dimer detection by the PMB tests. The positive reaction rate of hemoglobin and D-dimer is consistent with previous studies[5, 6]. Holtkötter et al.'s [5] study demonstrated a positive D-dimer reaction rate study demonstrated a positive D-dimer reaction rate in 70% of the postmortem samples (7 out of 10). For the MB samples, factors such as the day of the menstrual cycle on which the sample was collected could impact D-dimer detection. As expected, except for one sample, AM samples only react with hemoglobin, not D-dimer. This may be explained by certain hematologic conditions, such as deep vein thrombosis, which are known to be correlated with elevated D-dimer levels[3].

With respect to the Rapid Agglutination Assay, the values obtained are in agreement with previous literature[7], and demonstrated highest concentrations in PM and MB, while not being detectable or nearly undetectable in AM blood.

DNA profiling was successfully performed in all PM samples except for one degraded sample. This highlights the possibility of not only distinguishing PM blood, but also being able to identify the victim.

5. Conclusion

This study demonstrates that it is possible to efficiently use D-dimer lateral flow immunochromatographic tests to detect postmortem blood. This finding is supported by the highest concentrations of D-dimer being attributed to these samples. Detection of postmortem blood could be of utmost importance in crime scene reconstruction, e.g. when traces of blood are found but there is no corpse or when the body has been moved from the original crime scene. Additionally, this work points out the possibility to identify the victim through STR profiling of postmortem blood, which could also help with missing persons investigations.

6. Acknowledgments

The authors acknowledge Alexander Griberman and Dr. Christian Stadler for their revision of the manuscript and valuable comments. The authors would like to thank SERATEC® (SERATEC®, Göttingen, Germany) for providing the kits to carry out this work. This work was supported by the National Institutes of Health grant under Award Number R16-GM149358 (GTJ)

7. Conflict of interest statement

Dr. Gabriela Roca is an employee of SERATEC®.

8. References

[1] Sijen T, Harbison S: On the Identification of Body Fluids and Tissues: A Crucial Link in the Investigation and Solution of Crime. *Genes (Basel)* 2021, **12**(11).

[2] Riley RS, Gilbert AR, Dalton JB, Pai S, McPherson RA: Widely Used Types and Clinical Applications of D-Dimer Assay. *Lab Med* 2016, **47**(2):90-102.

[3] van der Hulle T, Tan M, den Exter PL, Mol GC, Iglesias del Sol A, van de Ree MA, Huisman MV, Klok FA: Selective D-dimer testing for the diagnosis of acute deep vein thrombosis: a validation study. *J Thromb Haemost* 2013, **11**(12):2184-2186.

[4] Wiewel-Verschueren S, Knol HM, Lisman T, Bogchelman DH, Kluin-Nelemans JC, van der Zee AG, Mulder AB, Meijer K: No increased systemic fibrinolysis in women with heavy menstrual bleeding. *J Thromb Haemost* 2014, **12**(9):1488-1493.

[5] Holtkotter H, Dias Filho CR, Schwender K, Stadler C, Vennemann M, Pacheco AC, Roca G: Forensic differentiation between peripheral and menstrual blood in cases of alleged sexual assault-validating an immunochromatographic multiplex assay for simultaneous detection of human hemoglobin and D-dimer. *Int J Legal Med* 2018, **132**(3):683-690.

[6] Konrad H, Hartung B, Poetsch M: (Un)Reliable detection of menstrual blood in forensic casework - evaluation of the Seratec(R) PMB test with mock samples. *Int J Legal Med* 2023.

[7] Sakurada K, Sakai I, Sekiguchi K, Shiraishi T, Ikegaya H, Yoshida K: Usefulness of a latex agglutination assay for FDP D-dimer to demonstrate the presence of postmortem blood. *Int J Legal Med* 2005, **119**(3):167-171.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1195-1202
DOI: https://dx.doi.org/10.15304/cc.2025.1869

The impact of temperature and humidity on the persistence, detection and recovery of semen

Tia Beveridge[a], Annalisa Durdle[a,b], Georgina E. Meakin[c], Roland A. H. van Oorschot[d,e], Bianca Szkuta[a*]

[a]*School of Life and Environmental Sciences, Deakin University, Geelong, Victoria, Australia*

[b]*Institute of Environmental Science and Research Limited, Private Bag 92021, Auckland 1142, New Zealand*

[c]*Centre for Forensic Science, University of Technology Sydney, Ultimo, New South Wales, Australia*

[d]*Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, Macleod, Victoria, Australia*

[e]*School of Agriculture, Biomedicine and Environment, La Trobe University, Bundoora, Victoria, Australia*

** Corresponding author at: e-mail:* b.szkuta@deakin.edu.au *Address: Locked bag 20000, Geelong, Vic, 3220. Ph: +61 3 5227 3132*

Abstract

Semen can be an important piece of evidence in criminal cases. To date, there has been little investigation into how environmental conditions impact the persistence, detection and recovery of semen and the associated DNA. Hence, this work aimed to investigate the impact of constant exposure to temperature and humidity combinations on the persistence, detection and recovery of semen that had been deposited onto cotton and left for 3 or 7 days. Semen (50 µL) was deposited onto swatches cut from new underwear and left in a climatic chamber at either 45°C or 15°C, with a relative humidity (RH) of 10 or 80%. These combinations were chosen to reflect Australian climates at different times of the year. Exposed samples were tested using acid phosphatase (AP) reagent, a lateral flow immunochromatographic test for the presence of prostate specific antigen (PSA), and hematoxylin and eosin staining to detect sperm. DNA was extracted and quantified. The results illustrated that when exposed to 45°C and 80% RH, AP activity is unable to be detected after 3 and 7 days of exposure, and DNA quantity is also negatively impacted. When left at the same temperature but at a RH of 10%, AP activity may be impacted after 7 days. PSA was detected in every instance, and sperm was observed on all slides except for one sample left at 15°C and 80% RH for 7 days. The results of this work will be useful for forensic investigators as it may help them to determine whether a sample may be suited for analysis when recovered from hot and/or

humid environments as well as their interpretation of test results given the case circumstances. The results may be used to highlight conditions of appropriate storage of evidentiary items in hot or humid climates.

Keywords
Semen, persistence, recovery, temperature, humidity, detection

1. Introduction

Semen can be an important piece of evidence in sexual assault cases and may be exposed to unfavorable environmental conditions for varying periods of time. There has been some research to date into how climatic conditions impact the persistence, detection and recovery of biological material [1,2]. Chen et al [1] observed DNA persistence using cell-free DNA on glass, polyester and cotton, and found that the quantities of DNA recovered declined significantly after 7 days of exposure to temperature conditions ranging between 15°C and 26°C with an average humidity of 64%, and that DNA generally persisted better under cooler conditions. Lee et al [2] studied the persistence of DNA in a Singapore context, and found that when samples of blood and touch DNA were left exposed to uncontrolled outdoor environmental conditions (temperature range: 23-35°C; RH: 50-99%), "highly variable persistence" was observed, compared to those left indoors under both controlled and uncontrolled conditions. All samples were exposed to conditions for up to 62 weeks for indoor and controlled conditions, and up to 4 weeks for outdoor conditions.

Understanding how semen and its components are impacted by climate and environmental conditions is critical to investigating their persistence, detection and recovery. One study observed the persistence of blood and semen specific mRNA and miRNA targets [3], where miRNA targets were able to be detected after 180 days, while mRNA was only detected up to 30 days. However, to date there has been no published research investigating how environmental conditions may impact other forensically relevant components of semen.

When detecting semen, commonly utilized tests include acid phosphatase reagent to screen for acid phosphatase enzymes, lateral flow immunochromatographic tests such as the p30® test (ABAcard® Diagnostics) for prostate specific antigens, and slide staining techniques, such as hematoxylin and eosin staining to visualize sperm. As crimes of a sexual nature may not get reported until days after the event, key evidence such as clothing may be exposed to weather conditions that

could potentially degrade these forensically relevant components of semen before they are collected for analysis. In this study, the aim was to investigate the impact of exposure to various temperature and humidity combinations for 3 or 7 days on the persistence, detection and recovery of semen.

2. Material studied, methods, techniques

Ethics approval (#2020-162) was received from the Deakin University Human Research Ethics Council prior to commencing laboratory work.

2.1 Substrate preparation

Benches and tools were cleaned with 1% sodium hypochlorite and deionized water. New, washed black cotton underwear (Anko, Kmart) was cut into 2.5 x 2.5 cm swatches. For each experiment, five swatches were taped to cleaned wire cooling racks (Kmart) with white masking tape and five taped to labelled blotting paper (COS Office Supplies).

2.2 Semen deposition

Semen was collected from one donor approximately 14 hours before deposition for each experiment. The donor did not ejaculate for at least 48 hours prior to donating samples. Each collected ejaculate (approx. 5 mL) was stored in a sterile container at 4°C until deposition. When required, semen was removed from the fridge and allowed to reach room temperature. The semen was vortexed for 10 seconds at high speed to homogenize the mixture. Semen (50 µL) was deposited on each cotton swatch.

2.3 Experimental conditions

2.3.1 Climate chamber

An ACS Compact DY(200T) climatic chamber was used. For each experiment, the climate chamber was set to the following conditions: Trial 1 - 45°C, 10% RH; Trial 2 - 45°C, 80% RH; Trial 3 - 15°C, 10% RH; Trial 4 - 15°C, 80% RH. All trials were left for 3 and 7 days. Trials 1 and 2 were conducted separately, while trials 3 and 4 were conducted at the same time. Wire racks with swatches were placed onto the middle shelf. After the experimental period, each group of swatches on wire racks was removed and placed on labelled blotting paper for source testing.

2.4 Controls

2.4.1 Experimental controls

For each experiment, three experimental controls were generated by taping swatches to blotting paper, depositing 50 µL of semen and leaving them in a biological safety cabinet (BSC) with a temperature and humidity data logger (HOBO, USA) for 3 or 7 days before source testing.

2.4.2 Positive and negative controls

Positive and negative controls were taken for each of the four experiments. The positive control consisted of 50 µL of semen deposited onto cotton at the time of source testing, while the negative control remained blank.

2.5 Source testing

2.5.1 Acid phosphatase

Filter paper (70 mm, Whatmann) was dampened with deionised water and firmly pressed onto each swatch for 5 seconds. The filter paper was removed, and lightly sprayed with acid phosphatase reagent (α-Naphtyl acid phosphatase, Fast Black K salt, sodium acetate, acetic acid and deionised water). The development of a purple color within 2 minutes was considered a positive reaction.

2.5.2 ABAcard® p30®

A 0.5cm x 0.5cm square was excised from the stained area using a scalpel (Swann-Morton No. 11) and placed into a labelled 1.5 mL Eppendorf tube. Sterile water (Fresenius Kabi) (650 µL) was added to the tube. The tube was vortexed for 5 seconds, and incubated at room temperature for one hour. The tube was vortexed for another 5 seconds and centrifuged for 3 minutes at 13,000 g. 200 µL of the extract was removed from the tube and added to a p30® test kit (ABAcard® Diagnostics). The results were read at 10 minutes. A positive result was defined as one line in both the control and test regions, while a negative result was defined as one line in the control section only. The remaining 450 µL of extract was frozen until required for DNA extraction.

2.5.3 Hematoxylin and eosin (H&E) staining

A 0.5cm x 0.5cm square was excised from the stained area using a scalpel (Swann-Morton No. 11) and placed into a labelled 1.5 mL Eppendorf tube, and

stored in the fridge until required. When required, 50 µL of sterile water was added to the tubes, and vortexed for 5 seconds. Ten microliters of extract was deposited onto a labelled microscope slide, and left overnight to dry. The stain was covered with hematoxylin and left for 15 minutes, then rinsed with tap water. The stain was covered with Scott's tap water, left for one minute, and rinsed with tap water. The stain was covered with eosin, left for one minute, and rinsed with tap water. The slides were rinsed with methanol. The slides were dipped 6 times in Histosolve, then 6 times in ethanol. The slides were left to dry, and then cover slips were added. Slides were viewed under 10 x and 40 x magnification, and scored as follows: -, no sperm detected; +, a few sperm, hard to find; 2+, some sperm in some fields, easy to find; 3+, many or some sperm in most fields; 4+, many sperm in every field.

2.6 DNA Analysis

Prior to extraction, the following reagents were prepared: Extraction buffer (1.21 g Tris, 3.72 g EDTA disodium salt, 5.84 g sodium chloride dissolved in 100 mL distilled water), Proteinase K solution (10mg/mL in dissolved water) and extraction digest buffer (extraction buffer and proteinase k). Lysis and wash buffers were prepared according to manufacturer's instructions. The 450 µL extracts were removed from the freezer and allowed to reach room temperature. One hundred microliters of extraction digest buffer was added to the sample, and incubated at 56°C for 30 minutes. The extract was transferred to a new, labelled 2 mL tube. DNA was manually extracted using the DNA IQ™ kit (Promega, Australia) to a final volume of 60 µL. Samples were quantified using the Quantifiler™ Trio Quantification Kit (Thermo Fisher Scientific, Australia) on an ABIPRISM® 7500 (Life Technologies, USA) and analyzed using HID Real-Time PCR Analysis Software.

3. Results

The results for each experiment are shown in Table 1. All controls produced expected results. For experimental control samples, temperature ranged from 16.5 to 25°C, and RH ranged from 35-62% for all experiments. As most 3- and 7-day experiments were conducted with different ejaculates of semen, the results of these experiments cannot be compared to one another. Instead, the results from each experiment are to be compared directly to the experimental controls generated for each of the respective experiments.

Table 1. Results for source testing and DNA quantities obtained for all experiments.

CC CONDITIONS	LENGTH OF EXPOSURE (DAYS)	AP		p30		H&E		DNA QUANTITY (ng) Min-Max (Med)	
		CC	EC	CC	EC	CC	EC	CC	EC
45°C, 10% RH	3 days	5/5	2/2[a]	5/5	2/2[a]	++ to ++++	+++[a]	46.5-727.4 (130.1)	27.2-36.1 (31.6)[a]
	7 days	4/5	2/3	5/5	3/3	++ to ++++	++ to +++	231.2-798.9 (693.6)	116.0 – 953.7 (913.7)
45°C, 80% RH	3 days	0/5	3/3	5/5	3/3	+ to +++	+++	3.6-105.7 (32.5)	30.5-379.4 (136.7)
	7 days	0/5	3/3	5/5	3/3	+ to +++	+++	0.06-32.8 (13.1)	165.9-1057.2 (1051.5)
15°C, 10% RH	3 days	5/5	3/3	5/5	3/3	++ to ++++	+ to +++	156.9-818.2 (621.8)	146.3-477.8 (214.1)
	7 days	5/5	3/3	5/5	3/3	++ to +++	+ to +++	84.9-673.2 (487.6)	146.4-214.9 (172.2)
15°C, 80% RH	3 days	5/5	3/3	5/5	3/3	+++ to ++++	+++	180.7-670.5 (415.4)	343.4-2060.4 (401.8)
	7 days	5/5	3/3	5/5	3/3	- to ++++	+++ to ++++	346.2-666.6 (635.7)	102.9-259.7 (161.4)

[a] Only two controls recorded.

4. Discussion

Acid phosphatase activity appeared to be impacted after samples were subjected to high temp and humidity (45°C, 80%) for both 3- and 7-days, though this was not the case for other temperature/humidity conditions or experimental controls. A considerably lower amount of DNA was recovered from both 3-day (3.6-105.7 ng, med. 32.5 ng) and 7-day (0.06-32.9 ng, med. 13.1 ng) samples that were exposed to high temp/humidity conditions compared to the respective experimental controls (3 days: 30.5-379.4 ng, med. 136.7 ng; 7 days: 165.9-1057.2 ng, med. 1051.5 ng). When samples were exposed to the same high temperature, but a lower humidity (10%) for 7 days, no acid phosphatase activity was observed in one sample. Further work is needed to investigate if this is due to the experimental conditions. When samples were exposed to lower temperatures, regardless of humidity, the median amount of DNA recovered from the climate chamber groups was higher than, or

within the same range as, their respective experimental controls (Table 1). The temperature of the room was between 16°C and 25°C during these experiments. These results suggest that DNA persistence and recovery may be impacted by higher temperatures, rather than humidity. These results are supported by Chen *et al.* [1], who recovered more DNA from samples subjected to cooler conditions and highlighted that DNA persistence was likely impacted more by temperature rather than humidity.

No impact on PSA detectability was observed, regardless of climate or control conditions, and at least one sperm was observed on all slides except for one sample left at 15°C and a RH of 80% for 7 days. DNA was still able to be recovered from this sample (644.66 ng). This suggests that PSA and sperm may be more resistant to the impacts of climate changes under the conditions investigated in this study.

When considering factors that impact DNA recovery, substrate type can also have an effect. Goray *et al*. [4] looked at secondary transfer of biological material on a range of porous and non-porous substrates and found that porous substrates tended to absorb the biological material, rather than transfer it when the substrate came into contact with another substrate, either porous or non-porous. In this study, only one porous substrate was used. Hence, some DNA may have been lost due to the absorption of the semen by the fabric in addition to being exposed to potentially harmful climate conditions. More studies are needed to investigate if substrate type, exposure time and environmental conditions, such as climate, have a significant impact on semen persistence.

5. Conclusion

Overall, the results of this study demonstrate that different components of semen may be affected by temperature, specifically high temperatures. They also highlight the need for additional studies exploring the impact that temperature and humidity have on the persistence, detection and recovery of semen. The results contribute to the currently limited knowledge of semen persistence and recovery when exposed to different climate conditions, and provide some understanding of when different components of semen may or may not be detected.

6. Acknowledgments

N/A

7. Conflict of interest statement

There are no conflicts to declare.

8. References

[1] Chen C, Pistono A, Ryan S. et al, The effect of climatic conditions on DNA persistence on glass, cotton and polyester, Forensic Sci. Int. Genet. Supp. Ser. 7 (2019) 274-276.

[2] Lee L. Y. C, Wong H. Y, Lee J. Y. et al, Persistence of DNA in the Singapore context. Int. J. Legal Med. 133 (2019) 1341-1349.

[3] Mayes C, Houston S, Seashols-Williams S. et al, The stability and persistence of blood and semen mRNA and miRNA targets for body fluid identification in environmentally challenged and laundered samples, Leg. Med. 38 (2019) 45-50.

[4] Goray M, Eken E, Mitchell R. J. et al, Secondary DNA transfer of biological substances under varying test conditions, Forensic Sci. Int. Genet. 4 (2019) 62-67.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1203-1208
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Far from home: genetic identification of Spanish Civil War victims

Miriam Baeta[1-3‡*], Caterina Raffone[1-3‡], Olatz Moreno-López[1,4], Lourdes Herrasti[4], Francisco Etxeberria[4,5], Belén Navarro-López[1-3], Marta Arroyo-Izaga[1,3,6], Iñigo Olalde[1,2,7], Marian Martínez De Pancorbo[1-3]

[1]*BIOMICs Research Group, Lascaray Research Center, University of the Basque Country UPV/EHU. Vitoria- Gasteiz (Spain).*

[2]*Department of Zoology and Animal Cellular Biology, Faculty of Pharmacy, University of the Basque Country UPV/EHU, Vitoria-Gasteiz (Spain).*

[3]*Bioaraba, BA04.03, Vitoria-Gasteiz (Spain).*

[4]*Department of Anthropology, Sociedad de Ciencias Aranzadi. Donostia-San Sebastián (Spain).*

[5]*Department of Legal Medicine and Forensics, University of the Basque Country UPV/EHU, Donostia-San Sebastián (Spain).*

[6]*Department of Pharmacy and Food Sciences, Faculty of Pharmacy, University of the Basque Country UPV/EHU, Vitoria-Gasteiz (Spain).*

[7]*Ikerbasque-Basque Foundation of Science, Bilbao (Spain).*

[‡]*Coauthors.*

**Corresponding author at: mbaetab@gmail.com*

Abstract

The genetic study of 70 victims of the Spanish Civil war (1936-1939) and posterior dictatorship (until 1970s) for identification purposes is reported. The victims, buried in graves in the cemetery of Orduña in Northern Spain (Basque Country), came from the central prison established in the locality. According to the official reports, more than two hundred males died in this prison because of the terrible conditions of insalubrity and hunger. Most of the victims were from southern regions of Spain and were sent to the north as part of the dispersion of prisoners far from home. Informative autosomal STR and Y-chromosome STR profiles were obtained from the skeletal remains of most of the victims, and were compared to those from relatives of 50 families. A total of 10 identifications have been reached at the moment. However, the search for new relatives of reference is continuing in order to identify the largest possible number of victims.

Keywords

Genetic identification, mass graves, Spanish Civil War, skeletal remains, ancient DNA

Introduction

During the Spanish Civil War (1936-1939) and the subsequent dictatorship (until the 1970s), an estimated 114,000 victims disappeared. Regrettably, 80 years after the outbreak of the conflict, the percentage of recovered or identified victims still represents a minor part of the totality. In fact, only approximately 9,000 victims have been recovered in the last eighteen years [1-3].

In 2014 and 2022, exhumations of the victims buried in graves in the cemetery of Orduña in North of Spain (Basque Country) were carried out. The victims came from the central prison established at that moment in the locality. According to the official reports, more than two hundred males died in this prison as a result of the terrible conditions of insalubrity and hunger. Most of the victims were from southern regions of Spain as they were sent to the north as part of the dispersion of prisoners far from home.

Figure 1. Exhumation of victims in the cemetery of Orduña (2022). Source: Sociedad de Ciencias Aranzadi.

The aim of the present study is the genetic analysis of skeletal remains of the 70 victims exhumed from the cemetery of Orduña, as well as alleged relatives, for identification purposes.

2. Material studied, methods, techniques

Skeletal remains of a total of 70 victims (mainly long bones, petrous bones or teeth) buried in graves in the cemetery of Orduña in North of Spain (Basque Country) were analyzed. According to official reports, the victims buried in the cemetery came from southern Spain; therefore, the search for relatives was possible based on documentary and historical reports, and institutionary support. Samples from alleged relatives belonging to 50 families were obtained; more concretely, buccal swabs were collected under informed consent, and in one case a skeletal remain from a relative was analyzed.

Prior to genetic analysis, anthropological studies were conducted to individualize the skeletal remains and obtained basic biological data (sex, age-at death, stature, cause of death, etc.)

Processing of the skeletal remains were performed as described in [2]. DNA extraction and quantification were carried out using the PrepFiler® Express BTA Forensic DNA Extraction Kit (Thermo Fisher Scientific, TFS, USA) in an Automate Express (TFS) and Quantifiler™ Trio DNA Quantification kit (TFS), respectively. DNA of buccal swabs from relatives was extracted using Gentra Puregene System (Qiagen, Germany) and quantified using NanoDrop spectrophotometer (TFS).

Autosomal STRs (aSTRs) and Amelogenin were studied using GlobalFiler™ IQC PCR Amplification Kit (TFS) and Y-STRs were analyzed using PowerPlex Y23 System (Promega Corporation, USA). Capillary electrophoresis was performed on 3500 Genetic Analyzer (TFS) and data analysis using Gene Mapper Software v.6.0 (TFS).

Comparison of genetic profiles obtained from remains and alleged relatives was carried out. Statistical calculation for aSTRs was performed with the statistical software Familias v.3.3 [4] or GFF v.2.56b. Y-STR haplotype frequency was assessed through YHRD (Y-STR Haplotype Reference Database) database.

In order to avoid contamination, recommendations suggested for work with ancient DNA (aDNA) were followed [5].

3. Results

In this study, the genetic analysis of skeletal remains of 70 victims of the Spanish Civil War and posterior dictatorship, buried in graves in the cemetery of Orduña in North of Spain (Basque Country), was carried out for identification purposes, as well as alleged relatives of 50 families.

A significant success in DNA profiling from skeletal remains recovered from Orduña graves was achieved. Informative profiles (≥ 12 markers) were obtained in 100% and 91% of the remains for aSTRs and Y-STRs, respectively. Figure 1 shows the success of DNA profiling for aSTRs and Y-STRs in skeletal remains, according to the number of markers obtained. For aSTRs, more than 17 markers were obtained in 70% of the remains and between 12 and 17 markers in the remaining 30%. For Y-STRs, in 58% of the remains more than 17 markers were obtained, in 33% between 12 and 17 markers and in the remaining 9% between 6 and 11 markers.

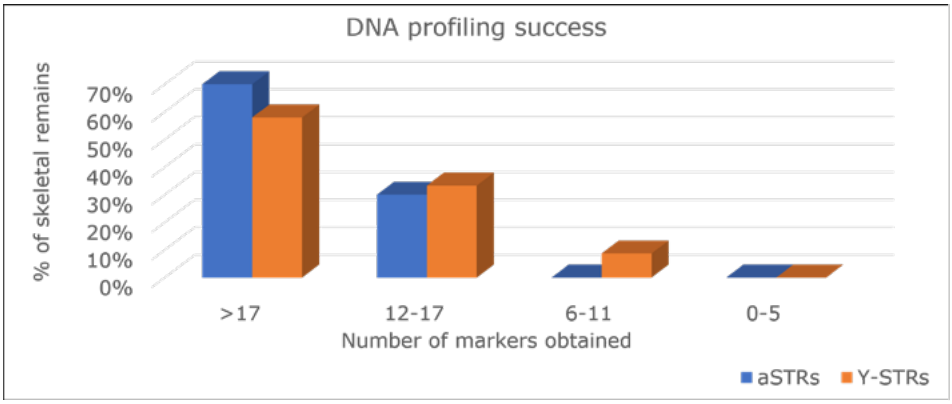


Figure 2. DNA profiling success for autosomal STRs (aSTRs) and Y-STRs in skeletal remains.

The genetic comparison performed between the skeletal remains and the presumed relatives allowed the identification of 10 victims. In more detail, five of the identifications were father/children cases, one paternal grandfather/ three grandchildren case, one maternal grandfather/three grandchildren case and one paternal uncle/nephew case. Two remains were identified as brothers, and their identity was stablished with one daughter and paternal relatives. LR values for aSTRs ranged from 2.44E+04 to 1.04E+10. In addition, patrilineal relationships were also supported by Y-STR haplotype matches between the remains and male relatives.

Other cases remain unresolved, despite the fact Y-STR haplotype matches have been determined between remains and potential relatives, as the match values obtained were below the established statistical threshold.

4. Discussion

The identification of victims of the Spanish Civil War and posterior dictatorship from exhumed skeletal remains is a challenging task, mainly due to the difficulty of obtaining genetic profiles from the remains. Skeletal remains typically contain low amounts of preserved DNA with variable degree of damage, as well as the presence of inhibitors, and endogenous or exogenous contamination [6].

This study has achieved significant success in DNA profiling of the skeletal remains of the 70 exhumed victims, as informative autosomal STR and Y-chromosome STR profiles were obtained in most cases. The overall success denotes the efficiency and sensibility of the DNA extraction and genotyping methods herein used.

The partial success rate in identification (10 victims) is mainly due to the lack of suitable relatives for genetic comparisons, a consequence of the long time elapsed since the end of the conflict, and the subsequent difficulty in finding first-degree relatives. The exhumation of close relatives emerges as an option of great interest as it would allow a more direct comparison with the victims. Furthermore, the dispersion of the victims along Spain in order to avoid their identification by acquaintances also complicates the search of relatives. In this sense, a national database including DNA profiles of all regions of Spain is expected to be very useful, especially in cases like this one where most of the victims came from southern regions of Spain and they were sent to the north as part of the dispersion of prisoners far from home.

Additionally, the study of additional markers (i.e. MPS panels) would be useful to resolve complex kinship relationships, where conventional markers are not discriminative enough.

5. Conclusion

In this study, the genetic analysis of skeletal remains of 70 victims of the Spanish Civil War and posterior dictatorship, exhumed from the cemetery of Orduña, have been successfully carried out. The efficiency of the methodology used herein has allowed a significant success in genetic profiling despite the intrinsic difficulties of the analysis skeletal samples. Victim identification has been limited by the lack of suitable relatives used for genetic comparisons. The search of more relatives as well as the analysis of additional genetic markers are necessary in order to increase the number of genetic identifications.

6. Acknowledgments

Funds were provided by Gogora- Institute for Memory, Coexistence and Human Rights of the Basque Government and by the Consolidated Research Group of the Basque Government, Microfluidics & Biomics Cluster UPV/EHU (IT-1271-19 and IT1633-22).

7. Conflict of interest statement

Authors declare no competing interest.

8. References

[1] F. Etxeberria, Exhumaciones contemporáneas en España: las fosas comunes de la Guerra Civil, Boletín Galego de Medicina Legal e Forense (2012) 18.

[2] M. Baeta, C. Nunez, S. Cardoso, L. Palencia-Madrid, L. Herrasti, F. Etxeberria, M.M. de Pancorbo, Digging up the recent Spanish memory: genetic identification of human remains from mass graves of the Spanish Civil War and posterior dictatorship, Forensic Sci. Int. Genet. 9 (2015) 1272–1279.

[3] L. Herrasti, N Marquez-Grant, F. Etxeberria, Spanish Civil War: The recovery and identification of combatants. Forensic Science International 320 (2021) 110706.

[4] D. Kling, A. O. Tillmar, T. Egeland, Familias 3-Extensions and new functionality. Forensic Science International: Genetics 13 (2014) 121-127.

[5] A. Cooper, H.N. Poinar, Ancient DNA do it right or not at all, Science 289 (2000) 1139.

[6] C. Raffone, M. Baeta, N. Lambacher, E. Granizo-Rodríguez, F. Etxeberria, MM. de Pancorbo. Intrinsic and extrinsic factors that may influence DNA preservation in skeletal remains: A review. Forensic Sci Int. 325 (2021) 110859.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1209-1213
DOI: https://dx.doi.org/10.15304/cc.2025.1869

LR values in the identification of victims of the Spanish Civil War (1936-1939). The experience of a big consortium

Eduardo Arroyo-Pardo[1*], Mar Costa-Roger[2], Claudia Lopes Gomes[1], Sara Palomo-Díez[1], Manuel Lozano García[1], Sara Bravo Gómez[1], Cristina Santos Pereira[3], Raquel Rasal Soteras[4], Nuria Bonet Martín[4], Assumpció Malgosa Morera[3], Eduardo Tizzano Ferrari[2], Enric Tartera Bieto[5], Ana María López-Parra[1]

[1]*Laboratory of Forensic and Population Genetics, Faculty of Medicine, Complutense University of Madrid, Spain.*

[2]*Area de Genética Clínica y Molecular, Hospital Vall d'Hebron, Barcelona, Spain.*

[3]*Laboratori d'ADN antic, Dep. Biologia Humana, Animal, Biologia Vegetal i Ecologia, Universitat Autónoma de Barcelona, Spain.*

[4]*Servei de Genomica, Dep. de Ciencies Experimentals i de la Salut, Universitat Pompeu i Fabra, Barcelona, Spain.*

[5]*Iltirta Arqueologia S.L., Lleida, Spain.*

** Corresponding author at: [1]Laboratory of Forensic and Population Genetics, Faculty of Medicine, Complutense University of Madrid, Spain. Phone: +34 913941576, E-mail: eduardoa@ucm.es.*

Abstract

We present data of the study promoted by the General Directorate of Democratic Memory of the Generalitat de Catalunya, between July 2017 and the present, to identify those who died as a result of the Spanish Civil War. A database of 3936 potential living descendants is available in the Clinical and Molecular Genetics Area of the Vall d'Hebron Hospital, Barcelona, Spain. Three centres, including Pompeu Fabra University, Autonomous University of Barcelona and Complutense University of Madrid, have provided a total of 579 (plus 62 with no results) genetic profiles of deceased people based on human remains excavated by the company Iltirta S.L. The identifications were made using the DVI module from Familias 3.2 software, first, and then LR value calculation was based on individual comparisons between suspected samples.

As of March 2024, there have been a total of 24 cases with identificative results, all of them belonging to males. In 6 cases, the identification was made through autosomal STRs with GlobalFiler PCR Amplification Kit (Thermo Fisher Scientific)(LR>>100000) and subsequent confirmation

with Yfiler Plus PCR Amplification Kit (Thermo Fisher Scientific), and in 2 cases with subsequent confirmation with mitochondrial haplotype. In addition, in 5 cases there could only be a patrilineal match probability by Y chromosome and in 1 case a matrilineal match by mitochondrial haplotype. Regarding kinship, in 11 cases daughters or sons were used, in 2 cases siblings, in 4 cases grandchildren, in 1 case great-grandchildren and in 6 cases nephews or nieces. Elevated LR values mostly correspond to parent-child comparisons that will decrease in frequency over time. Therefore, it is necessary to incorporate more polymorphism into the genetic profiles in order to be able to make comparisons between more distant relatives.

Keywords
Genetic identification, Spanish Civil War, Likelihood Ratio.

1. Introduction

The Spanish Civil War (1936-1939) resulted in numerous unidentified human remains. Recently, modern genotype technologies have proven to be a powerful tool for identifying these remains. Genetic results can add valuable information to data from physical anthropological origin that, normally, use to be insufficient [1]. Genetic profiles –both from autosomal, Y-chromosome and mtDNA origin- allow to test complex family pedigrees through probability and LR values in warfare cases [2,3]. In routine forensic practice, some commercial kits are used.

Hereby we present the identification data -obtained thanks to the aforementioned commercial kits- of the study promoted by the General Directorate of Democratic Memory of the Generalitat de Catalunya, between July 2017 and the present, to identify those who died as a result of the Spanish Civil War.

2. Material and methods

A database of 3936 potential living descendants is available in the Clinical and Molecular Genetics Area of the Vall d'Hebron Hospital, Barcelona, Spain. Three spanish centres, including Pompeu Fabra University, Autonomous University of Barcelona and Complutense University of Madrid, have provided a total of 579 genetic profiles of deceased people based on human remains excavated by the company Iltirta S.L. Identifications were made using the DVI module from Familias 3.2 software, first, and then LR value calculation was based on the Case-related DNA data module/individual comparisons between suspected samples.

Used kits were GlobalFiler PCR Amplification Kit and Yfiler Plus PCR Amplification Kit (Thermo Fisher Scientific). Mitochondrial DNA was typed through PowerSeq CRM Nested System kit (Promega). Living relatives were typed for mitocondrial DNA through Sanger sequencing (2 PCRs for positions 16021-16519 y 71-360).

3. Results

In the following table, results general results of the study, together with information of the samples, are presented (Table 1).

Table 1. UCM: Complutense University of Madrid, AUB/UPF: Autonomous University of Barcelona / Pompeu I Fabra University. Column Type of marker refers to the genetic system considered for identification.

Sample code	Identification	Location	Laboratory	Familiar relation sample / relatives	LR autosomal STRs	LR lineage markers	Type of marker
SOL68	2019	Cementiri vell del Soleràs	UCM	Maternal grandson / Paternal grandnephew-grandson	1178/6.41	4420	Autosomal STR and Y-STR
SOL71	2019	Cementiri vell del Soleràs	UCM	Maternal granddaughter	475269		Autosomal STR
SOL70	2019	Cementiri vell del Soleràs	UCM	Paternal great-grandson	0.90	262328	Y-STRs
SOL09	2019	Cementiri vell del Soleràs	UCM	Daugther	6411254		Autosomal STR
SOL28	2019	Cementiri vell del Soleràs	AUB/UPF	Paternal grandson	10.09	514160	Y-STRs
SOL69	2019	Cementiri vell del Soleràs	UCM	Paternal nephew	3.3	512945	Y-STRs
SOL85	2019	Cementiri vell del Soleràs	UCM	Paternal nephew	534.8	509808	Autosomal STR and Y-STR
CAST5	2019	Castellar del Vallès	AUB/UPF	Maternal nieces	-	13560	mtDNA
SOL06	2020	Cementiri vell del Soleràs	AUB/UPF	Paternal nephew	4.5	727129	Y-STRs
SOL01	2021	Cementiri vell del Soleràs	UCM	Sister	3.76E+08		Autosomal STR
REIG-1	2021	Els Reguers	AUB/UPF	Daughter	5.72E+07		Autosomal STR
SOL99	2021	Cementiri vell del Soleràs	UCM	Son	1.47E+07	800649	Autosomal STR and Y-STR
SOL102	2021	Cementiri vell del Soleràs	UCM	Son	5.44E+07	823567	Autosomal STR and Y-STR
MIR21	2018	Pernafeites	UCM	Daughter	2115614		Autosomal STR
MTR04	2021	Mas Torrenova / Moli C16	AUB/UPF	Daughter	1163926		Autosomal STR
MSM18	2021	Mas de Santa Magdalena	UCM	Son	4.20E+07	778747	Autosomal STR and Y-STR
MSM28	2022	Mas de Santa Magdalena	AUB/UPF	Daughter	4.37E+07		Autosomal STR
MSM127	2022	Mas de Santa Magdalena	UCM	Paternal niece and nephew	386625	568515	Autosomal STR and Y-STR
MSM136	2022	Mas de Santa Magdalena	UCM	Son	1.23E+08		Autosomal STR
SOR05	2020	Sorpe	UCM	Daughter	1812107		Autosomal STR
SOR07	2020	Sorpe	UCM	Maternal grandson / Paternal grandson / Paternal granddaughter	1.06E+23		Autosomal STR
SOR08	2022	Sorpe	UCM	Paternal nephew	50.91	674972	Y-STRs
UF41 / CR22	2023	Reus	AUB/UPF	Brother	1.28E+11	100	Autosomal STR and Y-STR
MSM149	2023	Mas de Santa Magdalena	AUB/UPF	Daughter and grandson	3.61E+13		Autosomal STR

As of March 2024, there have been a total of 24 identifications, all of them belonging to males. In 7 cases, the identification was made through autosomal STRs

with GlobalFiler PCR Amplification Kit (Thermo Fisher Scientific)(LR>>105) and subsequent confirmation with Yfiler Plus PCR Amplification Kit (Thermo Fisher Scientific). In addition, in 5 cases there could only be a patrilineal match probability by Y chromosome and in 1 case a matrilineal match by mitochondrial haplotype. Regarding kinship, in 11 cases daughters or sons were used, in 2 cases siblings, in 4 cases grandchildren, in 1 case great-grandchildren and in 6 cases nephews or nieces.

4. Discussion

The lack of non-genetic information about biological samples (historical, archaeological and anthropological data) decreases the number of targeted identifications, in which the LR is calculated between problem sample and a small group of living relatives.

Results seem to indicate that the LR decreases with kinship degree. Although Y-STR and mitochondrial DNA yields relevant LR values, highest LRs come from autosomal markers.

5. Conclusion

Elevated LR values mostly correspond to parent-child comparisons that will decrease in frequency over time. Therefore, it is necessary to incorporate more polymorphisms into the genetic profiles in order to be able to make comparisons between more distant relatives. Also, given that identifications may occur in the future, in the medium and even long term, it would be positive to add more polymorphisms in the current typing to obtain higher - and therefore conclusive - LR values in the future. Results seem to indicate that the LR decreases with kinship degree.

6. Acknowledgments

This project has been funded and promoted by General Directorate of Democratic Memory of the Generalitat de Catalunya.

7. Conflict of interest statement

The authors declare no conflict of interest.

8. References

[1] Manera-Scliar G., Hernández S., Martín-López M., Gomes. Difficulties in Kinship Analysis for Victims' Identification in Armed Conflicts. Genealogy 7,2 (2023): 31. https://doi.org/10.3390/genealogy7020031

[2] Thompson R., Zoppis S., McCord B. An overview of DNA typing methods for human identification: past, present, and future. Methods Mol. Biol. 830 (2012):3-16. doi.org/10.1007/978-1-61779-461-2_1. PMID: 22139649.

[3] Xu Y., Allen E., Wang L, Wen S. Identifying Human Remains from 20th Century Warfare: A State of the Field Essay. Front. Biosci. (Landmark Ed). 27(9) (2022):271. doi.org/10.31083/j.fbl2709271. PMID: 36224018.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1215-1221
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Identification of human remains in the Iraqi context

Mohammed Mashni Farhan[1,2], Ali Al-Janabi[1], *William H Goodwin[1]

[1]*University of Central Lancashire, Preston, United Kingdom.*
[2]*Medico-Legal Directorate, Baghdad, Iraq.*

** Corresponding author at: whgoodwin@uclan.ac.uk,*

Abstract

Iraq has experienced multiple conflicts over the last few decades with hundreds of thousands of deaths. Many of the victims have not been identified. The Medico Legal Directorate (MLD), which falls under the Ministry of Health, is authorized and mandated in law along with the Martyrs' Foundation, to carry out the exhumation and identification of human remains. While DNA analysis plays a critical part in most identifications the MLD has also developed and incorporates anthropological analysis, which is especially important in cases involving commingling.

Notable episodes of conflict and violence that have resulted in large numbers of deceased include: the Iraq-Iran war (1980-1988) with over 50,000 Iraqis still missing; the Al-Anfal campaign (1978-1988) where between 50,000 and 100,000 Kurds were killed; the Gulf War (1990-1991) with up to 50,000 Iraqi Service men killed; the Shia Uprise against Saddam regimen (1991) where between 400,000 and 1,000,000 were killed with most placed into mass graves, and the civil conflict (2005 -2008) that is still politically very sensitive. In its recent history between 2014 and 2017 ISIL (Islamic State of Iraq and the Levant) occupied approximately one third of the country and during this time committed many atrocities, including Camp Speicher where up to 1,700 air cadets were killed, Badoush Prison where over 600 Shia prisoners were executed and the persecution of the Yazidi population, where an estimated 5,000 males have been killed.

Each context presents its own challenges. Recovery of DNA from much of the skeletal material is difficult, with high levels of DNA degradation. Reference sample collection has been difficult in some contexts, with large numbers of relatives among the deceased and the challenges of tracing relatives following internal displacement and emigration. The security and location of some mass grave sites has also hampered the recovery of human remains.

Despite the challenges the MLD have been successful in identifying large numbers of individuals. Here we will discuss the incorporation of DNA in the identification process and provide more detail on the identification of the victims of the Yazidi population and Camp Speicher.

Keywords
Human identification, humanitarian, DVI, STR.

1. Introduction

In its recent history Iraq has suffered through international wars, civil wars, and terrorism. One result of the conflict is that there are more than 500,000 missing persons. Thousands of exhumed skeletal remains from over 300 mass graves dating to different period of times have been recovered and are awaiting identification; several hundred mass graves are still undiscovered or awaiting exhumation.

The Mass Graves Department was established in 2009 and is part of the MLD under the Ministry of Health. The DNA laboratory has been operational since 2013, with DNA analysis directly contributing to the identification of human remains since 2014.

While DNA analysis plays a critical part in most identifications the MLD also incorporates anthropological analysis, which is especially important in complex cases involving large numbers of missing and where commingled remains are recovered (Figure 1). The Identification Committee considers all lines of evidence before making an identification [1].

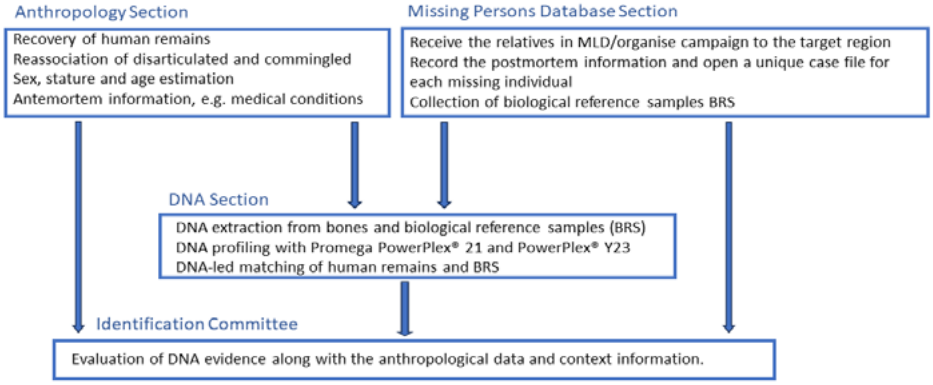


Figure 1. outline of identification process within the MLD

Human remains with very different levels of preservation are received and processed by the MLD ranging from samples only a few years old to those dating the Iran -Iraq war of the 1980s. While the processing of recent samples tends to be relatively routine even some samples from recent conflicts can be challenging because of environmental insult.

The Mass Graves Department has been actively processing samples for over ten years and has refined its extraction technique over this period. We present two case studies below that illustrate the challenges of the casework undertaken by the MLD and also highlights the potential to achieve identifications and return the human remains to the families of the missing.

2. Materials and methods

2.1. Material

Skeletal remains are exhumed from different mass graves in Iraq after a burial period ranging from a few years to over 30 years. Any soft tissue and external contamination on the bone samples are removed. Samples are cleaned using a Dremel or dental type rotary drill with a variety of sanding bits to accommodate different sizes and shapes of bone samples. The samples are then washed with sodium hypochlorite 5%, water, absolute ethanol and then dried for 24-48 h, depending on the sample condition.

All bone samples are pulverized using a Waring blender base, 1 L, model 7009 or 7011 and Waring blender cups, model SS110. Pulverized samples are stored at -20 °C until the time for DNA extraction.

2.2. DNA Extraction and DNA Profiling

Extraction for samples can vary slightly depending on the condition of the sample. Routine extractions have used semi-automated extraction with the Promega Maxwell ® 16. Typically, 0.5 g of ground material was incubated with 4 ml of bone incubation buffer for 24 h at 56 °C. The samples were concentrated using Amicon® Ultra-4 50 kDa columns. The resulting extracts of up to 250 ul were then purified using the DNA IQ™ Casework Pro Kit (Promega) following the manufacturer's instructions. An alternative method based on total demineralisation [2] was used in some case, typically when samples contained low amounts of DNA.

DNA extracts were quantitated using a 7500 Real Time PCR system with the HID Real-Time PCR Analysis (Life Technologies). The PowerQuant kit and the Pow-

erQuant analysis tool from Promega were used for quantification and the detection of PCR inhibitors. DNA amplification was carried out using a 7900 PCR system gold series (Life Technologies) with the PowerPlex 21 kit (Promega). Capillary electrophoresis and data analysis were undertaken using a 3500xL Genetic Analyzer and GeneMapper ID-X 1.5 (Thermo Fisher Scientific).

2.3. Matching and likelihood ratios

Matching of DNA profiles from reference and human remains was undertaken using GeneMarker v 2.7 (SoftGenetics) and likelihood ratios were calculated using Familias [3]. Likelihood thresholds are set depending on the context and in particular the likely number of victims [4, 5].

3. Results and Discussion

The identification projects undertaken by the MLD relate to different periods of conflict and varying scale. Below two case studies are presented.

3.1 Case study 1: Identification of members of Yazidi population

Most of Iraq's minority groups have historically resided in Nineveh (Northern Iraq) including the Yazidis, Assyrian and Chaldean Christians, Sabaean-Mandaeans, Turkmen, Shabak, and Kaka'i. ISIL specifically targeted these groups to "purify" the area of non-Islamic influences. Yazidis were singled out for particularly cruel treatment because ISIL members saw them as devil worshippers [6, 7].

According to the UN, at least 5,000 males were executed and 7,000 women and girls from the Shingal region (Sinjar mount and the surround area) were subjected to sexual slavery [7,8).

An identification program has been underway in Iraq to identify members of the Yazidi population that were killed during the ISIL occupation. To date, 786 human remains have been recovered and DNA has been extracted and profiled using Promega's PowerPlex® 21 STR kit and PowerPlex® Y23 kit. To provide samples for comparison an outreach programme has been undertaken to collect reference DNA samples along with other relevant antemortem data from relatives of missing individuals.

The programme faces numerous challenges such as location and recovery of human remains and locating biological relatives, many of whom are either internally displaced persons (IDPs) or have sought asylum in other countries. Additional

complexities in the case arise because of multiple victims within the same families. An example of one family is shown in Figure 2.

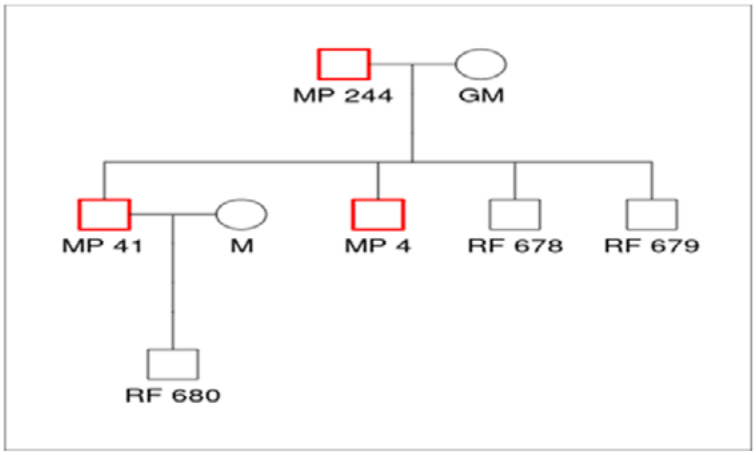


Figure 2. Pedigree of family with red squares for the missing persons (MP). GM: grandmother, M: mother, RF: reference sample.

For DNA-based identifications the Iraqi allele frequency database is used [9], but this is based on the Arabic population. A Yazidi allele frequency database was constructed (informed consent was provided by all individuals that provided a DNA sample). Exact tests comparing individual loci identified significantly different allele frequencies in 10 out of 15 loci that were compared.

Likelihood ratios were calculated for 105 sibling cases (i.e. sibling indexes) with all other relatives on the pedigrees excluded for the purpose of comparing the two reference databases. The median log10 value when using the Yazidi database was approximately one order of magnitude lower (5.06) than when using the Iraqi Arab allele frequency database (5.91), illustrating the importance of using the most relevant allele frequency database to reduce the potential for false matches.

So far 243 identifications have been made. The largest challenges remain the recovery of human remains and accessing victims' relatives. To date just over 2,000 of the estimated 5,000 victims have been registered as missing.

3.2 Case study 2: Camp Speicher

In June 2014 between approximately 1,700 young, unarmed Shia air cadets were seized and slaughtered in and around Tikrit, Iraq. ISIL officially claimed

responsibility for the well-reported massacre. Most victims were buried in mass graves with others dumped into the Tigris River. Following the removal of ISIL exhumations and recovery of bodies from the river started in the area.

In this context DNA was highly degraded even though the human remains were recovered after 3-4 years postmortem; however, full PowerPlex® 21 profiles could be generated for approximately 95% of the samples. At the end of 2023 1,139 bodies from the 1,700 victims had been identified and returned to their families. In addition to assisting with identifications, DNA, along with anthropological assessment has helped to reassociate the highly comingled remains of 48 individuals recovered from the Tigris River.

4. Conclusion

The Mass Graves Department within the MLD Iraq has been conducting human identification with the help of DNA for over 10 years. The contexts are complex, and the human remains often show elevated levels of degradation. However, despite the challenges 1000s of individuals have been identified and their remains returned to their families. Further work is ongoing at the MLD to improve the DNA recovery from degraded samples and examine the potential for SNP panels to support identification.

5. Acknowledgments

The MLD would like to thank all external agencies that have provided support to the Mass Graves Department and especially to the International Committee of the Red Cross (ICRC), the International Commission on Missing Persons (ICMP) and the United Nations Investigative Team to Promote Accountability for Crimes Committed by Da'esh/ISIL (UNITAD).

6. Conflict of interest statement

None

7. References

[1] M. Salado Puerto, D. Abboud, J.P. Baraybar, A. Carracedo, S. Fonseca, W. Goodwin, et al., The search process: Integrating the investigation and identification of missing and unidentified persons, Forensic Sci. Int. Synergy 3 (2021) 100154

[2] R. Huel, S. Amory, A. Bilić, S. Vidović, E. Jasaragić, T.J. Parsons, DNA extraction from aged skeletal samples for STR typing by capillary electrophoresis, in: A. Alonso (Ed), Methods Mol. Biol. 830 (2012) 185-98.

[3] Kling, D, Tillmar, A, Egeland, T, Familias 3-Extensions and new functionality, Forensic Sci. Int. Genet. 13 (2014) 121-127.

[4] C. Brenner, Some Mathematical Problems in the DNA Identification of Victims in the 2004 Tsunami and Similar Mass Fatalities, Forensic Sci. Int. 157 (2006) 172-180.

[5] C. Peel, W. Goodwin Theoretical value of the recommended expanded European Standard Set of STR loci for the identification of human remains. Med. Sci. Law. 52 (2012) 162-168.

[6] V. Cetorelli, S. Ashraph, Sareta, A demographic documentation of ISIS's attack on the Yazidi village of Kocho, LSE Middle East Centre, London, 2019.

[7] D. van Zoonen, K. Wirya, The Yazidis. *Perceptions of Reconciliation and Conflict,* Middle East Research Institute, Erbil, 2017.

[8] P. Nicolaus, S. Yuce, Sex-Slavery: One Aspect of the Yezidi Genocide. *Iran and the Caucasus,* 21 (2017) 196-229.

[9] M. M. Farhan, S. Hadi, A. Iyengar, W. Goodwin (2016) Population genetic data for 20 autosomal STR loci in an Iraqi Arab population: Application to the identification of human remains. Forensic Sci. Int. Genet. 25 (2016) e10-e11.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1223-1225
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Monitoring success rates of touch DNA samples in cases of volume crime in a high-throughput German police lab

Jens Söchtig[1*], Cadja Lassen[1], Christian Vidal[1]

[1]*State Criminal Police Office Lower Saxony, Forensic Science Institute, Hannover, Germany*

** Corresponding author at: State Criminal Police Office Lower Saxony, Forensic Science Institute, Schützenstr. 25, 30161 Hannover, Germany, Tel. +49 511 98735180*

Abstract
Our study focuses on the challenges and success rates of analysing touch DNA in volume property crimes like burglaries and vehicle break-ins. Despite advancements in DNA typing technology, touch DNA analysis remains complex due to factors such as limited DNA quantity or complex mixtures. However, law enforcement agencies often request touch DNA analysis under pressure to solve these crimes. To assess its effectiveness and to interpret the value of trace DNA evidence, the Forensic Science Institute of the State Criminal Police Office Lower Saxony monitored success rates since 2017. Up until now, the analysis of 2671 contact traces from 1802 cases resulted in 65 DNA profiles, of which 19 matched in the German DNA database. The study's findings have been annually shared with law enforcement agencies which led to a decline in requests for certain types of contact traces.

Keywords
DNA analysis, Touch DNA, Latent DNA, Contact stain

1. Introduction

Volume property crime cases (e.g. burglary and vehicle break-ins) are characterized by a high number of contact traces considered as low-template touch DNA. These contact traces form a substantial part of the workload for the high-throughput forensic DNA laboratory of the Lower Saxony State Office of Criminal Investigation [1]. Decades of research and development improved sensitivity of DNA typing methods tremendously, but still, analysis of touch DNA presents with

numerous challenges [2]. Success rates of contact traces remain low due to limited DNA quantity / quality, complex mixture interpretation, and the absence of DNA reference samples. However, facing social and political pressure to increase success rates in cases of volume property crime, law enforcement agencies frequently request the analysis of contact traces aiming to provide the criminal justice system with objective and reliable evidence [3]. Aiming to depict a realistic valuation of touch DNA evidence in volume crime investigation, we started to examine our contact trace success rates.

2. Material studied, methods, techniques

In 2017, we have implemented a continuous process of monitoring results of analysing touch DNA samples in real cases of volume property crime. The entire workflow of contact trace processing follows accredited routine protocols and the concluding success rate data is collected manually. For burglary, we are individually focusing on pry marks, (glove) swipe marks, and drill holes. Looking at cases of vehicle break-in, we concentrate on infotainment systems (e.g. contact traces from various infotainment connectors and plugs) as well as screws.

3. Results

From 2017 to 2024, 2671 contact traces belonging to 1802 cases have been analysed and evaluated. Altogether, 65 DNA profiles were sent to the German DNA database (DAD) of the Federal Police Department (*Bundeskriminalamt*) resulting in 19 hits. Seeing this low likelihood of success, requests for pry marks, (glove) swipe marks, and car-infotainment plugs are typically rejected (except for well-justified reasons). In comparison, success rates for drill holes (often accompanied by saliva stains) are superior justifying their analysis.

4. Discussion and Conclusion

The police intranet of Lower Saxony is provided annually with the success rates of each contact trace type. This data collation offers objective statistics to better evaluate the potential outcome of contact DNA evidence in volume crime investigation

for the investigative phase in the criminal justice system of Lower Saxony. In accordance with that, recent trends show a drop in this type of investigation requests.

Difficulties were encountered during the data collation. Each case had to be searched manually, which is a time-consuming task. Given the high workloads with routine casework, this is given less priority. To further address the challenges of manual data collection, we are planning to automate success rate estimations using our Laboratory Information Management System (LIMS).

As another consequence of this study, special training of police officers can help to improve success of crime investigations.

5. Acknowledgments

The authors wish to thank the technicians of the DNA laboratory of the Forensic Science Institute of the State Criminal Police Office Lower Saxony

6. Conflict of interest statement

The authors declare that they have no competing interests.

7. References

[1] Reuss E, Kauferstein S, Zehner R: Sicherung und Auswertung von latenten DNA Spuren im Bereich der Eigentumskriminalität. Rechtsmedizin 2008

[2] Mapes A, Kloosterman A, de Poot C: DNA in the Criminal Justice System: The DNA Success Story in Perspective. J Forensic Sci. 2015

[3] Raymond J, van Oorschot R, Gunn P, Walsh S, Roux C: Trace DNA success rates relating to volume crime offences, FSI:Gen. 2009

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1227-1233
DOI:

The role of NGS mitogenomes and DNA phenotyping in the forensic investigation: a casework of newborn abandoned in a dumpster

Carla Bini[1*], Sara Amurri[1], Giulia Fazio[1], Filomena Melchionda[2], Chiara Turchi[2], Cristina Giuliani [3], Donata Luiselli[4], Susi Pelotti[1]

[1]*Department of Medical and Surgical Sciences, Section of Legal Medicine, University of Bologna,, Italy.*

[2]*Department of Biomedical Sciences and Public Health, Marche Polytechnic University, Ancona, Italy.*

[3]*Department of Biological, Geological and Environmental Sciences, Laboratory of Molecular Anthropology & Centre for Genome Biology, University of Bologna, Bologna Italy.*

[4]*Laboratory of Ancient DNA, Department of Cultural Heritage, University of Bologna, Ravenna, Italy.*

** Corresponding author at: via Irnerio 49, 40126 Bologna, +39 051 2088354, carla.bini@unibo.it (C. Bini).*

Abstract

Inside a large bag collected in a dumpster, a newborn was found, noticed due to the cries heard by passersby. When discovered in May 2013, the female infant, still alive, was wrapped in several towels soiled with widespread bloodstains. Additionally, the placenta and umbilical cord were also recovered near the baby. At first, the placenta samples from the mother's side were analysed to genotype autosomal DNA profile and mtDNA haplotype using the Sanger method for the HVS-I and HVS-II regions to determine the potential ancestral background of the unknown mother. The predicted haplogroup based on mutations in the control region was H27, which was predominantly found in northern Europe. In addition, biogeographic ancestry and DNA phenotyping analyses were performed using the Precision ID Ancestry Panel which targets 165 ancestry informative autosomal Single Nucleotide Polymorphisms (SNPs) and the 41 HIrisPlex-S markers respectively, on the Ion GeneStudio S5 Prime System. The results showed the mother's sample to have a predicted predominant European ancestry with Southwest Asian admixture. To increase the informativity, a massively parallel sequencing (MPS) experiment was performed using the Ion PGM system to analyse the entire mitogenome. The results revealed new mutations in the coding regions, leading to the assignment of a different haplogroup (H2a5), exclusively found in the Franco-Cantabrian area.

The HIrisPlex-S web tool predicted dark brown hair with a light shade, green eye colour, and an intermediate skin phenotype.
This case is an example of the impact of the development of new forensic markers and technologies considering that DNA-based inference of biogeographic ancestry should include ancestry-informative autosomal, Y-chromosomal and mitochondrial DNA markers to infer an individual's subcontinental ancestry. Combined DNA analysis of external visible characteristics (EVCs) and biogeographic ancestry can increase the information yield of a criminal investigation.

Keywords
Forensic genetics, Human identification, NGS technologies, Mitogenome, Forensic DNA Phenotyping, Biogeographic Ancestry Prediction

1. Introduction

In forensic casework, lineage markers are highly valued as they can provide information on the biogeographic ancestry of donors of crime scene trace samples, particularly in situations where there are few or no investigative leads or where eyewitness testimony is unavailable [1]. Analysis of sequence polymorphisms of the hypervariable mitochondrial DNA (mtDNA) regions HVI and HVII remains an important tool for this purpose, as mtDNA haplotypes are strongly correlated with geographic origin. By comparing mtDNA sequences with databases such as EMPOP, it is possible to identify specific haplotypes and trace an individual's ancestry. However, these hypervariable regions contain only 25% of mitochondrial DNA variants, and therefore most of the discrimination power of mitochondrial DNA lies in the coding region [2]. To increase the informativity of mtDNA, it is therefore possible to analyse the entire mitogenome using massively parallel sequencing (MPS) [3]. In addition, recent advances in Next-Generation Sequencing (NGS) allow the prediction of Biogeographical Ancestry (BGA) using autosomal Ancestry Informative Markers (AIMs) [4] and Externally Visible Characteristics (EVC) [5] through the analysis of informative autosomal SNPs, providing investigators with a more comprehensive genetic profile.

In May 2013, a placenta was found in a dumpster in the city of Bologna, along with the umbilical cord and several towels soiled with extensive blood stains, next to an abandoned newborn baby who was still alive.

Forensic DNA analyses were performed to identify the genetic profiles of the foetus and mother and to investigate the maternal biogeographical ancestry and EVCs. The analysis focused on advanced SNP genotyping and NGS technologies to provide key insights into the individual's biogeographical ancestry and physical traits, greatly enhancing the forensic investigation with useful information.

2. Materials and methods

DNA from samples of the placenta on the mother's side and from the umbilical cord of the foetus were isolated by QIAmp DNA Investigator Kit (Qiagen, Hilden, Germany) and amplified with the PowerPlex® ESX 17 System (Promega); the HVS-I and HVS-II regions were amplified using primers L15997/H16401 and L00029/H00408, then sequenced using the BigDyeTerminator v1.1 kit (Applera) to determine the mtDNA haplotype; biogeographic ancestry prediction and DNA phenotyping analysis were performed using the Precision ID Ancestry Panel which targets 165 ancestry informative autosomal SNPs and the 41 HIrisPlex-S markers respectively [6], on the Ion GeneStudio S5 Prime System; the EVC data were evaluated using the "HirisPlex Prediction Webtool" (https://hirisplex.erasmusmc.nl) and the methodology proposed by Pośpiech et al. [7] for inconclusive results; a massive parallel sequencing (MPS) experiment was conducted using the Ion PGM System to analysis the entire mitogenome.

3. Results

The genetic profile of the newborn was easily genotyped from the umbilical cord sample, allowing a single-source female profile of the mother to be reconstructed from the maternal side of the placenta.
Based on the analysis of mutations in the HVI and HVII control regions on the maternal sample, the predicted haplogroup was identified as H27, which is predominantly found in Northern Europe. The results of the ancestry prediction analysis showed that the maternal sample had a predicted predominant European ancestry (66.75%) with Southwest Asian admixture (31.50%) (Figure 1.a). FDP analysis using the HIrisPlex-S web tool [6] predicted dark brown hair with a light shade and an intermediate skin phenotype. The inconclusive result for eye colour was then evaluated according to Pośpiech et al [7] and a prediction of a green eye colour phenotype was reported (Figure 1.b). The mitogenome analysis confirmed the mutations in the D-loop region detected by Sanger sequencing but also revealed new mutations in the coding regions. These additional findings were crucial for accurate assignment of the H2a5 haplogroup. This subgroup is indigenous to the Basque Country and is found only there, with the highest concentration (17%) in the province of Guipuzcoa. It represents a Franco-Cantabrian lineage that has been present in the Basque Country since prehistoric times (Figure 2) [8].

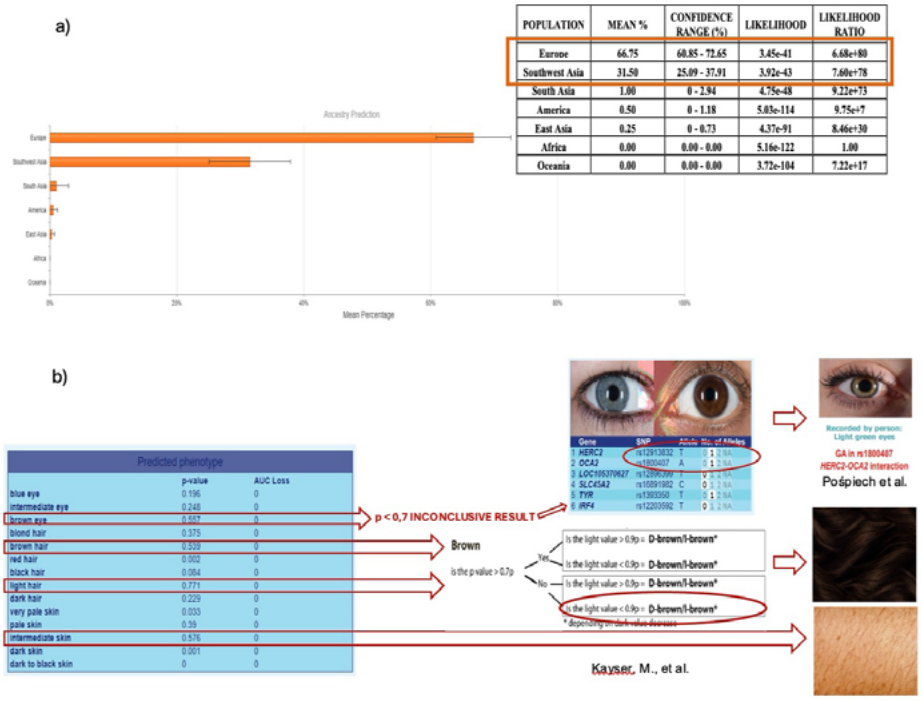


Figure 1. Results - a) Ancestry Prediction Admixture results- Bar chart showing the admixture prediction for the sample analysed using the Precision ID Ancestry Panel v1.0. The table shows the admixture prediction with confidence range and population likelihoods with variability estimates based on bootstrapping analysis. b) Forensic DNA Phenotyping results- Flowchart illustrating the process of predicting eye, hair and skin colour probabilities based on p-values obtained using the HIrisPlex-S prediction webtool(https://hirisplex.erasmusmc.nl) and using the methodology proposed by Pośpiech et al. for inconclusive eye colour results.

POPULATION	MEAN %	CONFIDENCE RANGE (%)	LIKELIHOOD	LIKELIHOOD RATIO
Europe	66.75	60.85 - 72.65	3.45e-41	6.68e+80
Southwest Asia	31.50	25.09 - 37.91	3.92e-43	7.60e+78
South Asia	1.00	0 - 2.94	4.75e-48	9.22e+73
America	0.50	0 - 1.18	5.03e-114	9.75e+7
East Asia	0.25	0 - 0.73	4.37e-91	8.46e+30
Africa	0.00	0.00 - 0.00	5.16e-122	1.00
Oceania	0.00	0.00 - 0.00	3.72e-104	7.22e+17

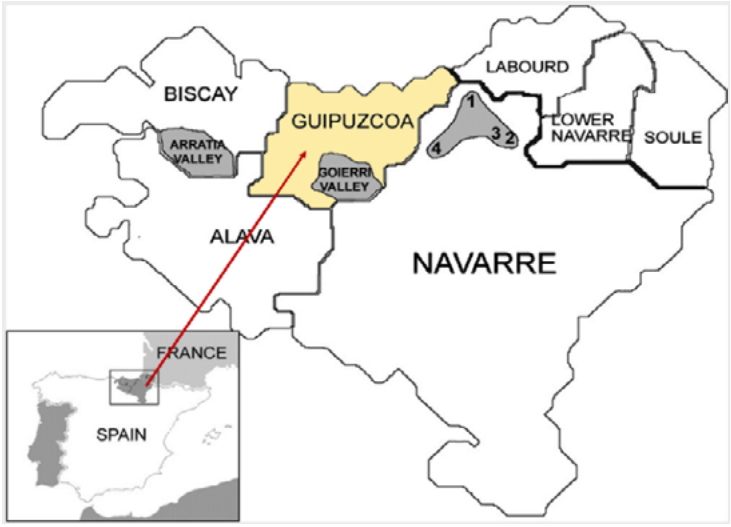


Figure 2. Province of Guipúzcoa in the Basque Country, northern Spain, where the highest concentration of the H2a5 haplogroup is found.

4. Discussion

The results of the biogeographical ancestry (BGA) analyses reveal inconsistencies that warrant further investigation. Specifically, while the whole mitogenome analysis identified the H2a5 haplogroup, which is found exclusively in the Franco-Cantabrian area, the Precision ID Ancestry Panel results showed a broader and more heterogeneous geographical area that includes both Europe and Southwest Asia. This suggests that mitogenome analysis may provide greater resolution than the autosomal SNPs used to infer the ancestral origins of an individual. The variation in geographical resolution can be attributed to the distinct inheritance patterns: mitochondrial DNA (mtDNA) traces only the maternal lineage, whereas autosomal SNPs reflect both maternal and paternal contributions [9]. Furthermore, although STR markers are well established as primary tools for personal identification, their use as AIMs is limited due to the presence of identical STR alleles in different populations resulting as a result of recurrent mutations [1,10]. This complicates the interpretation of genetic data and highlights the importance of careful marker selection in ancestry studies, as well as the need for the integration of different markers.

5. Conclusion

This case study provides an example of the significant impact that advances in forensic markers and technologies have had on the field. The results show that accurate DNA-based inference of biogeographic ancestry requires the incorporation of ancestry-informative autosomal, Y-chromosomal and mitochondrial DNA (mtDNA) markers. Furthermore, the integration of DNA analysis of externally visible features with biogeographic ancestry, particularly the role of NGS mitogenomes. significantly enhances the informational yield of criminal investigations, thereby improving the overall effectiveness of forensic analysis.

6. Conflict of interest statement

The authors declare that they have no conflict of interest.

7. References

[1] Y. Wen, J. Liu, Y. Su, et al., Forensic biogeographical ancestry inference: recent insights and current trends, *Genes Genom*. 45 (2023) 1229–1238.

[2] M. Stoljarova, et al., Whole mitochondrial genome genetic diversity in an Estonian population sample, *Int. J. Legal Med.* 130 (2016) 67–71.

[3] J.R. Connell, M.C. Benton, R.A. Lea, et al., Evaluating the suitability of current mitochondrial DNA interpretation guidelines for multigenerational whole mitochondrial genome comparisons, *J. Forensic Sci.* 67 (2022) 1766–1775.

[4] M. Gabriel, C. Buchanan-Wright, A. Kumar, et al., Analytical improvements in biogeographic ancestry inference. Presented at the 28th Congress of the International Society for Forensic Genetics (ISFG) (2019) Prague, Czech Republic, Thermo Fisher Scientific.

[5] M. Kayser, W. Branicki, W. Parson, et al., Recent advances in forensic DNA phenotyping of appearance, ancestry and age, *Forensic Sci. Int. Genet*. 65 (2023) 102870.

[6] K. Breslin, B. Wills, A. Ralf, et al., HIrisPlex-S system for eye, hair, and skin color prediction from DNA: massively parallel sequencing solutions for two common forensically used platforms, *Forensic Sci. Int. Genet*. 43 (2019) 102152.

[7] E. Pośpiech, J. Draus-Barini, T. Kupiec, et al., Gene-gene interactions contribute to eye colour variation in humans, *J. Hum. Genet*. 56 (2011) 447–455.

[8] A. Gómez-Carballa, A. Olivieri, D.M. Behar, et al., Genetic continuity in the Franco-Cantabrian region: new clues from autochthonous mitogenomes, *PLoS One,*7(3) (2012) e32851.

[9] D. McNevin, Forensic inference of biogeographical ancestry from genotype: the Genetic Ancestry Lab, *WIREs Forensic Sci.* 2 (2020) e1356.

[10] E. Alladio, C. Della Rocca, F. Barni, et al., A multivariate statistical approach for the estimation of the ethnic origin of unknown genetic profiles in forensic genetics, *Forensic Sci. Int. Genet*. 45 (2020) 102209.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1235-1241
DOI: https://dx.doi.org/10.15304/cc.2025.1869

MPS Proficiency testing results of the GHEP-ISFG program on STR, Y-STR, and mtDNA markers

Dr. Koro Fernández*[1] Dr Ana Mosquera-Miguel[2] Dr Walther Parson[3], Dr Mishel Stephenson[4*]

[1]*National Institute of Toxicology and Forensic Sciences*

[2]*Forensic Genetics Unit, Institute of Forensic Sciences, Universidade de Santiago de Compostela*

[3]*Institute of Legal Medicine, Medical University of Innsbruck, Forensic Science Program, The Pennsylvania State University, University Park*

[4]*Forensic Anthropology Foundation of Guatemala*

** Corresponding author at: National Institute of Toxicology and Forensic Sciences, C/José Echegaray, 4, 28232 Las Rozas de Madrid. (Spain).coro.fernandez@justicia.es*

Abstract

The increased number and types of genetic markers that can be analyzed simultaneously, the higher throughput of samples and the improvement of results in cases of degraded DNA, among others, are the main reasons to the expansion of the Massive Parallel Sequencing (MPS) technology in the field of forensic genetics. Once this technique has been validated, a forensic DNA laboratory has to implement quality controls and monitor that it is working properly. In 2023, a MPS Proficiency testing, using the samples of the basic level of the EIADN Exercise coordinated by the National Institute of Toxicology and Forensic Sciences (INTCF) and organized within the Spanish and Portuguese Speaking Group of the ISFG was offered to kinship and forensic laboratories. Results of the analysis of three reference items and two forensic items were compared to those obtained by capillary electrophoresis. Twelve laboratories took part analyzing STR markers and additionally seven performed analysis of the mitochondrial DNA control region. In general results were quite good. Regarding autosomal STRs, there was one laboratory that reported one repeat less in all genotypes for the D6S474 system with respect to the other laboratories. Moreover, it was observed that not all laboratories reported results using the same strand direction. As for the Y chromosome markers, no consensus was reached for the marker DYS612 due to different nomenclatures used. Finally, errors in mitochondrial DNA analyses were due to items swap, low quality sequence, transcription error or due to wrong interpretation of the AC repeat in the HV3 region. The development of MPS Proficiency testing is essential to evaluate the performance

of laboratories, identify inter-laboratory differences and initiate actions within the forensic community for improvement of the methodology, software analysis and nomenclature.

Keywords
Massive Parallel Sequencing, Forensic Proficiency Testing, Quality Assurance System, Forensic DNA, Spanish and Portuguese Speaking Group of the ISFG (GHEP-ISFG).

1. Introduction

The increased number and types of genetic markers that can be analysed simultaneously, the higher throughput of samples and the improvement of results in cases of degraded DNA, among others, are the main reasons to the expansion of the Massive Parallel Sequencing (MPS) technology in the field of forensic genetics. Once validated this technique, a forensic DNA laboratory has to implement quality controls and monitor that it is working properly. In 2023, a MPS Proficiency testing, using the samples of the basic level of the EIADN 31 Exercise coordinated by the National Institute of Toxicology and Forensic Sciences (INTCF) and organized within the Spanish and Portuguese Speaking Group of the ISFG, was offered to kinship and forensic laboratories. The aim of this Exercise was to provide laboratories with an external tool to assess and monitor their performance. In this work we present the results of the analysis of three reference items and two forensic items which were compared to those obtained by capillary electrophoresis (CE). STR markers (autosomal, Chromosome Y and Chromosome X) and mitochondrial DNA were analyzed.

2. Material studied, methods, techniques

2.1. Participants

A total of 12 participants from European and American kinship and forensic laboratories took part in the Exercise. They were requested to use the markers they employ in their laboratories: 12 laboratories analysed Autosomal STR (A-STR), seven Y-STR, five X-STR and seven mitochondrial DNA (mtDNA).

2.2. Items and requirements

2.2.1 Kinship module:

M1: 50 µl blood from a male onto an Ahlstrom-Munksjö GenCollect card.

M2: 120 µl saliva from a female onto a Blood Stain Storage System (BSSS), Whatman card.

M3: 120 µl saliva from a female on a cotton swab.

2.2.2 Forensic Module.

M4: 50 µl of a 1:2(v/v) mixture of blood from a male (donor of item M1) and saliva from a female on a piece of wipe.

M5: hair from a female.

2.2.3 Requirements

Genetic analysis of the three reference items and the two forensic items obtained with MPS technology was requested. Results had to be reported including both CE allele calls and the corresponding sequence string. Formats for STR sequences: Bracketed repeat regions and/or STR naming according to ISFG nomenclature [1,2] and for mitochondrial DNA haplotypes: medical or forensic nomenclature was requested.

2.3. Assessment

STR: A general assessment of results was performed by comparison with those obtained with the capillary electrophoresis (CE) analyses for those markers in common in the EIADN 31 (2023) PT. This value was reported but no individual assessment was performed, as such, so the laboratory had to perform a self-evaluation. Mixtures of X chromosome STRs markers were not evaluated.

Mitochondrial DNA: Each of the mtDNA regions analyzed (HV1, HV2 and HV3) were evaluated separately. Assigned values were obtained by consensus. The length variants present in the homopolymeric tracts were not assessed. This includes insertions or deletions that may appear at positions: 16193, 309, 455, 463 and 573. Mixtures of mtDNA were not evaluated.

2.4. Methodology

2.4.1 Panels

STR kits: Five laboratories used Precision ID GlobalFiler NGS STR Panel (Thermofisher (TFS)), four ForenSeq DNA Signature Prep kit Mix A (Verogen), one ForenSeq DNA Signature Prep kit Mix B (Verogen) and two ForenSeq MainstAY kit (Verogen).

Mitochondrial DNA kits: Three laboratories used the Precision ID mtDNA Control Region Panel (TFS), two ForenSeq mtDNA Control Region Kit (Verogen), one Precision ID mtDNA Whole genome Panel (TFS), and one Early Access Mito Kit v.1 (Applied Biosystems).

Data analysis tools: Different software were used for the analysis of raw data reading, allele calling and final analysis depending on the panels and platforms used (Table 1).

Table 1. Software used for the analysis of MPS sequence data

	n labs	raw data	allele calling	analysis
STR	3	TSS	HID GP	CV
	1	TSS	TVC	CV
	1	TSS	STRait Razor	Manual
	7	UAS	UAS	UAS
Mt DNA	2	TSS	HID GP	CV
	1	TSS	TVC	CV
	1	TSS	CV	CV
	1	TSS	IGV	Manual
	2	UAS	UAS	UAS

TSS: Torrent Server Suite (TFS)
UAS: Universal Analysis Software (Verogen)
HID GP: HID Genotyper plugin (TFS)
TVC: Torrent Variant Caller (TFS)
CV: Converge (TFS)

3. Results

Kinship module: In general results were quite good for all the markers analyzed.

A-STR markers: Errors were concentrated mainly on one laboratory, which indicated that they had had problems with the kit used. The lab had sequencing errors at four markers for M1, four for M2 and six markers for M3.

One laboratory reported one repeat less in all genotypes for the D6S474 system with respect to the other laboratories, based on the Forensic Sequence STRucture Guide v.6 *[3]*, as indicated.

Several laboratories identified the presence of SNPs in the flanking region which showed no gain in alleles numbers.

Regarding isoalleles within the STR sequence, they were observed at locus D5S818 in M1, M2 and M3.

Y-STR markers: Two laboratories presented discrepancies in M1. Errors were attributed to failure to obtain results for three markers and to the presence of an artifact in marker DYS438 (12-13 instead of 12).

No consensus was reached for locus DYS612.

Only one laboratory reported the presence of SNPs in the flanking region.

X-STR markers: results were concordant among participants and consistent with the CE analysis. The same lab as for Y-STR reported the presence of SNPs in the flanking region.

Mitochondrial DNA: In general, high read depth was obtained. Only two errors were detected: one due to an unreported variant and the other due to a nomenclature error.

Forensic module

A-STR markers: A consensus was reached for all markers except for D6S474 and D12ATA63. Errors were concentrated on two laboratories, and they were mainly due to partial drop-outs of some of the alleles or due to failure to obtain results in some markers.

Several laboratories identified isoalleles at loci D21S11, D5S818, D1S1656 and D12S391 with sequence variation. SNPs in the flanking regions were reported for markers D7S820 and D13S317.

As in the kinship module, one laboratory reported one repeat less for the D6S474 system with respect to the other laboratories.

Y-STR markers: As the male contributor in M4 was M1, the same discrepancies were found for markers DYS438 and DYS612. The same laboratory as for M1 reported the presence of SNPs in the flanking region.

Mitochondrial DNA in M5 (hair): Results were very good. Only one error was reported by one laboratory (309DEL), however C-Insertions at 309 were ignored for the evaluation. The deletion of C309 is a signature variant in some haplogroups and therefore significant. (73G 150T 228A 251A 263G **309DEL** 315.1C 16086C 16192T 16270T 16304C)

Regarding nomenclature, it was observed that not all laboratories report results using the same strand direction, for example D1S1656: [TAGA]15 [TAGG]1 vs [CCTA]1 [TCTA]15.

A-STR Six out of twelve laboratories reported the STR sequences using both the bracketed repeat region format and the ISFG nomenclature. Two, only the ISFG and four did not reported the sequence obtained with any of the nomenclatures proposed.

Y-STR Two out of seven reported the sequences obtained using the bracketed repeat region format.

X-STR One of five laboratories reported results using the bracketed repeat region format.

4. Discussion

Regarding STR analysis, results were in general correct when comparing them to those obtained with capillary electrophoresis data, except for marker D6S474. When typing this marker with the Qiagen Investigator HDplex in the CE analysis the resulting allele calls were one repeat unit less. Only one out of six laboratories reported the same alleles as for the CE data as they use the $[TAGA]_n$ TGA $[TAGA]_n$ motif following the recommendations of the FSSG v6 *[3]*. The rest of participants applied the $[AGAT]_n$ $[GATA]_n$ motif. In consequence, each motif results in a different length-based genotype: This issue has been reviewed and discussed in literature [4].

Different ways of reporting DYS612 alleles among laboratories has also led to genotype discordancy. Five out of seven laboratories analyzing this marker considered the sequence [TCT]n CCT [TCT]n, instead of [CCT]n CTT [TCT]n CCT [TCT]n, reporting six less repeats. In ForenSeq UAS Software the first two triplets ([CCT]n CTT) are not taken into account; on the contrary STRBase establishes this nomenclature [CCT]a [CTT]b[TCT]c [CCT]d [TCT]. By the time this Exercise was performed, Parson, Ballard, Budowle et al. [1] recommend not including initial repetitions [CCT]a [CTT]b. New ISFG recommendations issued in 2024 [2] conclude to include the entire region in the repeat count.

5. Conclusion

The most significant errors were due to the use of different motifs for reporting, causing ambiguities in allelic lengths. It is necessary that laboratories follow the current ISFG recommendations in order to harmonize genotype reporting such as implementation of the minimum reporting range recommended or the use of universal bioinformatic software for STR sequence data analysis. In addition, vendors´ software should be updated to implement the ISFG minimum range STR sequence data.

6. Acknowledgments

The coordinator of this exercise would like to express her gratitude to the participating laboratories for their contribution and to the advisers (co-authors) who performed the revision and assessment of the results. Finally, she wishes to thanks to the Spanish and Portuguese Speaking Group of the ISFG for their support.

7. Conflict of interest statement

All authors declare that they have no conflicts of interest.

8. References

[1] Parson, W. Ballard, D. Budowle, B. *et al. Massively parallel sequencing of forensic STRs: considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements, Forensic Sci. Int. Genet. 22 (2016) 54–63.*

[2] Gettings, K.B. Bodner, M. Borsuk, A. *et al. Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on short tandem repeat sequence nomenclature. Forensic Sci. Int. Genet. 68 (2024) 1029-46.*

[4] Bodner, M. Ballard. D, Borsuk, L.A. et al. Harmonizing the forensic nomenclature for STR loci D6S474 and DYS612. Forensic Sci. Int. Genet. 70 (2024) 103012.

"Web references":

[3] Forensic Sequence STRucture Guide (FSSG) v6 https://strider.online/nomenclature

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1243-1248
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Prüm implementation in the UK: An example of success through collaboration

Andrew Revoir[1*], Holly Lancaster[1], Katherine Bache[1], Shazia Khan[1]

[1]*Forensic Services, Metropolitan Police Service, London, United Kingdom*

**Corresponding author: Andrew.Revoir@met.police.uk*

Abstract
After conducting a successful pilot and receiving subsequent parliamentary approvals, the Metropolitan Police Service (MPS) on behalf of UK policing, went live with its Step 1 Prüm DNA service in July 2019, initially connecting with Austria followed quickly by Germany, the Netherlands and then Spain. Since then connections to further European countries have been established such that currently the UK MPS Prüm DNA team is now routinely sharing DNA biometric data with 27 countries and performing scientific verification of all matches prior to reporting to UK forces and engaging with the UK National Crime Agency (NCA) as the UK Step 2 partner for exchange of demographics. This review will consider the journey from pilot to go-live as well as highlighting the positive impact the implementation has had on the MPS, the UK and also partners across Europe since implementation as a crime fighting intelligence tool.

Keywords
United Kingdom, Prüm, DNA, Forensic, Database, Intelligence,

1. Introduction

The initial Prüm agreement was signed in 2005 and at this point the UK opted out. However, realising the potential benefits of biometric data sharing with EU Member States, a pilot was conducted by the MPS on behalf of the Home Office in 2015. Results of the initial pilot in 2015 were encouraging and lead to the Secretary of State for the Home Department recommending that the UK opt in to Prüm cooperation as detailed in [1].

Following parlimentary approval to participate in Prüm exchanges, the MPS was selected by the UK Home Office as the Step 1 partner for Prüm and the UK National Crime Agency (NCA) as the Step 2 partner. Step 1 concerns the automated searching of relevant DNA profiles (crime stains and subjects) and the scientific verification of any matches whereas Step 2 relates to the gatekeeping of requests for associated hit information such as requests for sharing of subject demographics in relation to generated hits.

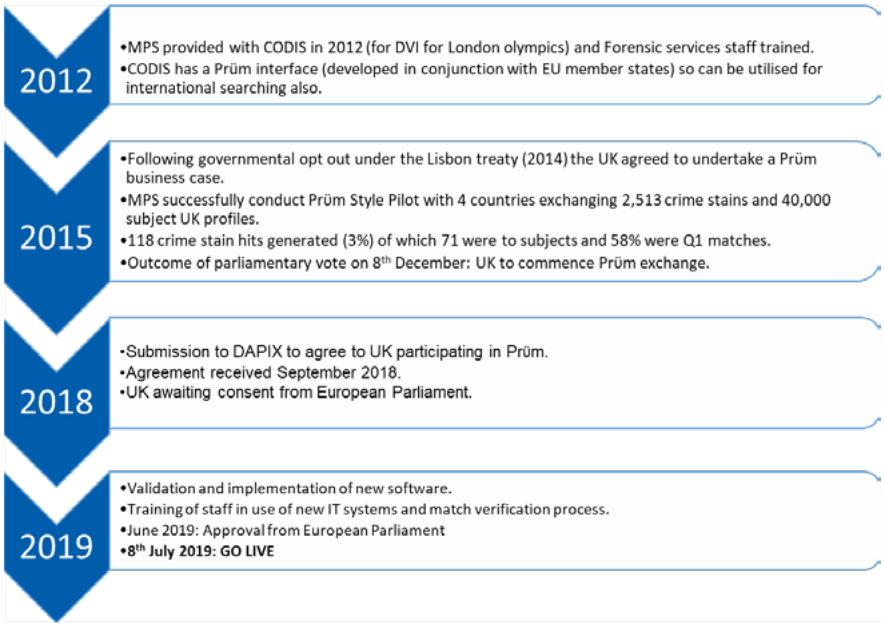


Figure 1. Timeline for MPS implementation of Prüm including successful completion of pilot, parlimentary appoval to re-join the Prüm mechanism and then Go-Live in 2019.

Upon selection as the step 1 partner the MPS Prüm DNA team then initiated a project to scale up the process in preparation for exchanging all applicable UK DNA data with future EU connections. This involved implementing novel software and processes to enable eligible DNA data from the UK National DNA database (NDNAD) to be made available for searching with EU partners and maintained to ensure it reflected the current status of records on the UK NDNAD. Key milestones from pilot through to go-live are detailed in figure 1. Figure 2 displays the process that was implemented for data maintenance at go-live.

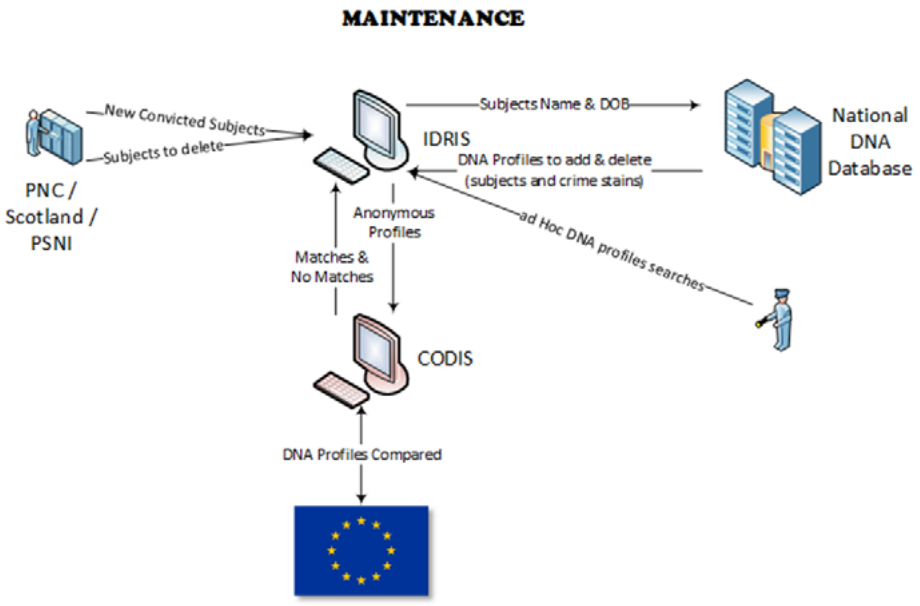


Figure 2. Diagram illustrating the exchange of various information between agencies in the UK enabling the maintanance of the UK Prüm DNA database such that it refelcts the staus of records on the UK NDNAD and ensures the correct DNA profile data is made available for searching. IDRIS (International DNA Record Information System) is a novel information management system designed to facilitate the management of information from different agencies and systems within the UK and with Prüm participants

Since implementation the Prüm DNA Service has now been integrated into the UK as a national Service and as part of this any crime stains unmatched to the ND-NAD are automatically searched against the 27 European connected countries as a matter of course. The UK NDNAD is already a well established and effective DNA intelligence tool but the implementation of Prüm DNA searching in the UK has proven to be an important supplement to this.

2. Material studied, methods, technique

Once the new go-live process, including supporting systems, had been designed and configured, new staff were recruited and trained to enable the delivery of the new service. Staff numbers were based on the anticipated number of matches based

on the initial planned exchanges with other Prüm DNA participants and those countries' database sizes. Since go-live in 2019, the UK has conducted testing and established connections with 27 Prüm DNA partner countries. The countries the UK is connected with and the date of connection is displayed in figure 3. In addition to training staff in system maintenance, a number of DNA hit reporting staff were trained in Prüm DNA match verification for all outgoing matches (for UK forces) as well as incoming matches (requested by Prüm DNA partner countries). This includes review of Q4 matches for potential discordance or designation errors.

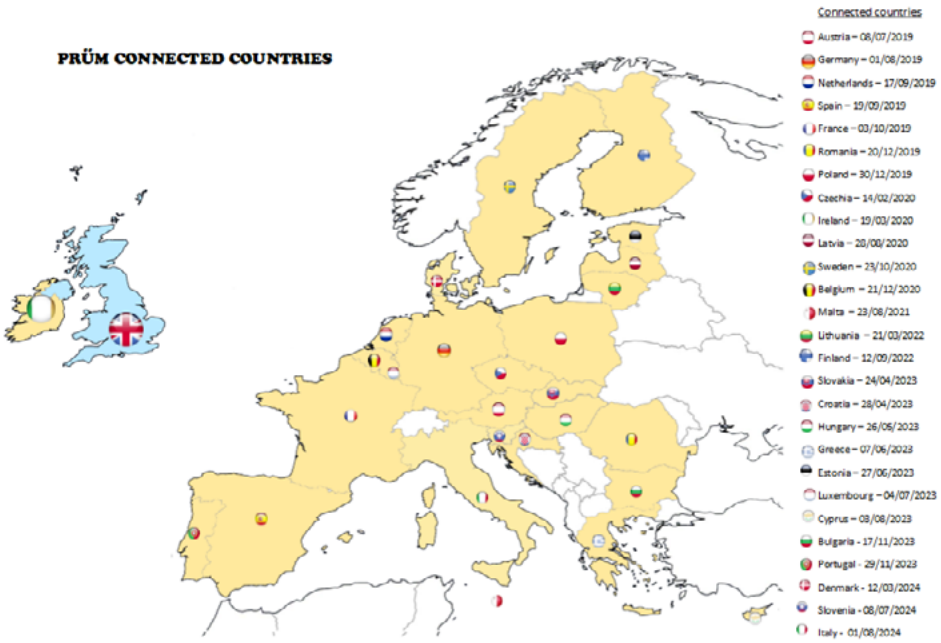


Figure 3. Map including the location of all 27 Prüm DNA partner countries that the UK is connected to. The figure also has the date of connection to all countries including the initial go-live connection with Austria on the 08/07/2019. The last country connected to was Italy on the 01/08/24.

3. Results

Since go-live a large number of autosomal STR profile matches have been generated from the UK DNA data exchanged through the Prüm mechanism which has been facilitated by the large size of the UK NDNAD. This has led to the generation of intelligence through large numbers of UK crime scene profile to member state subject profile matches (and vice versa), UK crime scene profile to member state crime scene profile matches as well as UK subject profile to member state subject profile matches. At the time of writing there is 6.5 million UK subject STR profiles available to search for all Prüm DNA connected countries as well as over a quarter of a million UK crime stain profiles. In addition in terms of matches so far the MPS DNA Prüm step 1 team has sent a total of 14,870 match notifications to either UK police force contacts or the NCA. Of these 2559 related to crime scene to crime scene matches and 5392 related to UK crime stain matches to member state subject profile records.

4. Discussion

Since the Prüm DNA service was implemented by the MPS a large number of matches have been generated which has benefitted both UK police forces as well as collegues in Europe. Not only has this enabled forces to gain intelligence by identifiying suspects through crime scene to subject matches but it has also allowed the identification of crime patterns through crime scene to crime scene matches which assists in detecting and preventing organised cross border crime.

The UK already has an effective domestic DNA database with a high hit rate (chance that a crime scene DNA profile, once loaded onto NDNAD, matches against a subject DNA profile stored on the NDNAD) consistently above 60% with the lowest value of 61.9% and the highest value of 66.9% in the last 10 years [2]. By implementing the Prüm DNA service which applies to all crime stains unmatched to the NDNAD it is now possible to automatically further search unmatched UK DNA profiles against a large number of DNA STR profiles held across 27 connected countries across Europe. This can provide intelligence through DNA matches which was previously not possible for a large number of unmatched DNA profiles on the NDNAD. It also provides a valuable additional DNA intelligence database tool to supplement the NDNAD and other more specialized options such as familial searching of data held within the NDNAD and Interpol searches. The Prüm DNA

service also has the added benefit that it can be utilized for all crime types unlike familial searching and Interpol searches which are generally restricted to serious crimes only in the UK.

5. Conclusion

The MPS Prüm DNA team, working in conjunction with partners in the UK and Europe, succeeded in completing a proof-of-concept pilot and then implementing the Step 1 DNA Prüm Service. This is now embedded in the UK as part of the range of national biometric tools available providing intelligence in investigations ranging from serious crime (such as murder and rapes) to volume crime (burglary and vehicle theft). The Service is also supporting connected European countries by providing access to the UK's large DNA profile dataset, thereby supporting them in criminal investigations outside the UK.

6. Acknowledgments

All members of the MPS Prüm DNA team who helped implement and currently maintain the service for the UK as well as collegues at other government organisations such as the NCA and FINDS. Also UK Forensic Service Providers and partners in other Prüm DNA connected European countries who also enable this important service to function effectively.

7. Conflict of interest statement

None of the above authors has any conflict of interest in relation to this publication.

8. References

[1] UK Home Office. Prüm Business and implementation case. (2015). Accessible online at https://www.gov.uk/government/publications/prum-business-and-implementation-case. Accessed 22/08/2024.

[2] Forensic Information Databases annual report 2022 to 2023. (2024) at *https://www.gov.uk/government/publications/forensic-information-databases-annual-report-2022-to-2023/forensic-information-databases-annual-report-2022-to-2023-accessible*. Accessed 16/09/2024.

30th CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Universidade de Santiago de Compostela, 2025, pp. 1249-1253
DOI: https://dx.doi.org/10.15304/cc.2025.1869

Towards a Common Body of Knowledge for Forensic Genetics: the Most Valuable Publications List and the INTERPOL DNA Reviews

John M. Butler[1*]

[1]*National Institute of Standards and Technology, Special Programs Office, USA*

** Corresponding author at: National Institute of Standards and Technology, 100 Bureau Drive, Mail Stop 6021, Gaithersburg, MD 20899-6021, USA, john.butler@nist.gov*

Abstract

An effort to identify and describe some of the most valuable publications (MVPs) in forensic genetics has been made with the initial 2021 MVP list containing almost 500 informative publications across 26 topic categories. MVP lists were revised in 2022 and 2024 with fewer citations. These lists have been informed by two INTERPOL reviews on forensic biology and forensic DNA typing published in 2020 and 2023.

Keywords

Forensic biology, forensic DNA, forensic genetics, training resources.

1. Introduction

Effective training and continuing education are crucial to keep up with evolving forensic genetics technologies and applications. Forensic laboratories invest in the continuing education of their staff. Universities with forensic science programs seek to prepare their students to be future contributing practitioners and researchers. Stakeholders in the criminal justice system (e.g., law enforcement personnel, lawyers, and judges) using DNA results also benefit from regular training and continuing education to understand capabilities and limitations of methods and

practices. Almost 100 books on forensic genetics have been published since 1990 including six by the author [1].

An agreed-upon common body of knowledge in forensic genetics benefits practitioners, students, and stakeholders [2]. Thousands of publications in dozens of peer-reviewed journals exist on the topic of forensic genetics – and information in this field grows with each passing year due to many active research efforts. This ever-growing body of scientific literature becomes increasingly challenging to monitor, much less incorporate into forensic laboratory training programs. Regular updates on important advances and resources with helpful articles in specific areas of interest to forensic DNA casework can assist DNA technical leaders and analysts.

2. Creating MVP Lists

Over the past several decades, INTERPOL has organized a triennial symposium for forensic science managers to discuss important advances. Review articles on select forensic disciplines are prepared by invited experts and presented to those in attendance. The last two sets of reviews have been published as special issues in an open-access format with the journal *Forensic Science International: Synergy* [3]. These reviews have covered the topics of fingermarks and other body impressions; paint, tape, and glass evidence; fibers and textiles; fire debris and fire investigation; firearm examination; toxicology; questioned documents; forensic management; drug analysis; gunshot residue; forensic video analysis; explosives; digital evidence; and forensic biology and DNA typing.

The author was invited by the INTERPOL symposium organizers to prepare the two most recent reviews on forensic genetics (biology and DNA), where 235 articles from 35 journals across 12 topics were discussed in the 2016-2019 review [4] and 773 articles from 96 journals across 15 topics with the 2019-2022 review [5]. Efforts in gathering hundreds of articles and studying the literature for the field inspired a desire to create a list of impactful information – the most valuable publications (MVPs).

Input on the MVPs came through activities of the Organization of Scientific Area Committees (OSAC) for Forensic Science and the Scientific Working Group on DNA Analysis Methods (SWGDAM). The initial MVP list contained almost 500 informative publications across 26 topic categories and was introduced in a 2021 workshop at the American Academy of Forensic Sciences (AAFS). Additional MVP lists have evolved and been somewhat more focused as other presentations were given and feedback was received. These reference lists build upon information cited

in the July 2020 SWGDAM Training Guidelines [6] and were revised in 2022 and 2024 in conjunction with presentations at various conferences [7]. The categories and the MVP reference lists created so far can be improved upon, and suggestions for additional references and categories are welcome.

3. MVP Topics and Organization

Table 1 lists the 26 topics and the number of articles in each category or group for the three MVP lists created so far in 2021, 2022, and 2024. Within each group (listed arbitrarily from A to Z), a #1 MVP article was selected, which is often an ISFG DNA Commission article, and then followed by reference citations, which are not ranked further, defined by date in ascending order with the most recent publications at the end of each category. This letter and number system (e.g., A1, B3, W7) provides a simple method to locate specific articles in each MVP list and enables opportunities for expansion as the literature grows. Although some articles could logically appear under multiple categories, no duplicate listings were used.

Table 1. The most valuable publications (MVP) topics covered in 26 groups (categories) along with the number of articles present in each category for the three lists created in 2021, 2022, and 2024.

Group (Category)	**Most Valuable Publications Topic(s) Covered** https://strbase.nist.gov/Information/Most_Valuable_Publications	**480** (2021)	**85** (2022)	**92** (2024)
A	Plain Language Guides to Forensic DNA Analysis	4	2	3
B	Serology and Body Fluid Identification	24	3	3
C	Collection and Storage of Biological Material	25	2	2
D	DNA Extraction/Purification, Differential Extraction	18	2	2
E	DNA Quantitation, Degraded DNA	10	2	2
F	PCR Amplification, Inhibition, and Artifacts	13	3	3
G	Capillary Electrophoresis Separation and Detection	12	2	2
H	Assessing Sample Suitability & Complexity, Low-Template	7	2	2
I	Estimating the Number of Contributors	12	4	4
J	Data Interpretation, Mixture Deconvolution, and Interlaboratory Studies	12	4	5
K	Interpretation: Binary Approaches (CPI, RMP, LR)	11	5	5
L	Interpretation: Probabilistic Genotyping Software	44	4	6

M	Report Writing and Technical Review	8	4	4
N	Court Testimony, Communication, Juror Comprehension	22	5	5
O	Autosomal STR Markers and Kits	29	2	2
P	Mitochondrial DNA Testing	11	3	3
Q	Y-Chromosome and X-Chromosome Testing	17	4	4
R	DNA Databases and Investigative Genetic Genealogy	14	3	3
S	Statistical Analysis	11	2	2
T	Population Genetics	11	2	2
U	DNA Phenotyping (Ancestry, Appearance, Age)	24	2	2
V	New Technologies (Rapid DNA, Massively Parallel Sequencing)	35	5	5
W	DNA Transfer and Activity Level Reporting	57	8	10
X	Non-Human DNA Testing	15	2	2
Y	Method Validation, Quality Control, and Human Factors	23	5	5
Z	General Forensic Science Topics	11	3	4

4. Conclusion

The NIST STRBase website now includes past and present MVP lists as well as associated presentations [7]. The MVP lists are not intended to be additive (i.e., using 2021 + 2022 + 2024), and some of the best references are retained over time as new ones are added and some removed. Interested users are encouraged to use the most recent version – with the ability to access previous versions as desired. Note that some years (e.g., 2023) may not have an updated MVP list.

Ideally, in the future a group of informed researchers and practitioners would build upon this effort and create revised MVP lists every year to evolve with advances in the field. Of course, any selection of best articles is going to be subjective and not comprehensive.

5. Acknowledgments and Disclaimer

Funding was provided by the NIST Special Programs Office Forensic Science Foundation Studies Program. Several individuals have influenced this effort including the following: for input on the initial MVP list (Phil Danielson and an OSAC Forensic Biology literature task group), for input and co-presenting in an AAFS 2021 workshop (Robin Cotton, Mecki Prinz, Charlotte Word, Amy Brodeur, and Teresa Cheromcha), and for creation of the STRBase webpage (Lisa Borsuk and Pete Vallone).

Points of view in this document are those of the author and do not necessarily represent the official position or policies of the National Institute of Standards and Technology. Certain commercial software, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

6. Conflict of interest statement

There are no conflicts of interest to report (the author does not receive royalties for his books).

7. References and Notes

[1] The author's *Forensic DNA Typing* books were published in 2001, 2005, 2009, 2011, 2014, and 2022.

[2] Based on deliberations and discussions of NIST team members and Resource Group in connection with the scientific foundation review on DNA Mixture Interpretation (NISTIR 8351-draft, Appendix 2); see also *Forensic DNA Interpretation and Human Factors: Improving Practice Through a Systems Approach*, Chapter 9 "Education, Training, and Professional Credentialing", pp. 241-274. https://doi.org/10.6028/NIST.IR.8503

[3] Interpol special issues in *Forensic Science International: Synergy*:
(2016-2019) https://www.sciencedirect.com/special-issue/10QWNF78TVR
(2019-2022) https://www.sciencedirect.com/special-issue/109JJLCFP14

[4] J.M. Butler, and S. Willis, INTERPOL review of forensic biology and forensic DNA typing 2016-2019, *Forensic Sci. Int.: Synergy* 2 (2020) 352-367. https://doi.org/10.1016/j.fsisyn.2019.12.002

[5] J.M. Butler, Recent advances in forensic biology and forensic DNA typing: INTERPOL review 2019-2022, *Forensic Sci. Int.: Synergy* 6 (2023) 100311. https://doi.org/10.1016/j.fsisyn.2022.100311

[6] SWGDAM Training Guidelines (2020) Available at https://www.swgdam.org/publications

[7] See https://strbase.nist.gov/Information/Most_Valuable_Publications

C U R S O S E C O N G R E S O S
Nº 278

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA