Methods in Molecular Biology 1420

Springer Protocols

William Goodwin *Editor*

Forensic DNA Typing Protocols

Second Edition

EXTRAS ONLINE

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Forensic DNA Typing Protocols

Second Edition

Edited by

William Goodwin

School of Forensic and Applied Sciences, University of Central Lancashire, Preston, Lancashire, UK

Humana Press

Editor
William Goodwin
School of Forensic and Applied Sciences
University of Central Lancashire
Preston, Lancashire, UK

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-3595-6 ISBN 978-1-4939-3597-0 (eBook)
DOI 10.1007/978-1-4939-3597-0

Library of Congress Control Number: 2016941950

1st edition: 2005

Printed on acid-free paper

This Humana Press imprint is published by Springer Nature
The registered company is Springer Science+Business Media LLC New York

Preface

Forensic DNA analysis has evolved almost beyond recognition in the last 30 years. Following its inception in 1985, the first ten years saw rapid changes in technology with short tandem repeats (STRs) replacing variable number tandem repeats (VNTRs) as the DNA marker of choice. From the mid-1990s to the present day, the changes in technology have been less dramatic and STRs remain the main tool utilized by forensic laboratories around the world, albeit with incremental improvements in reagents, equipment and practice. These changes have made DNA profiling more powerful, sensitive, and resistant to PCR inhibitors.

While STRs remain the mainstay of forensic analysis, several tools have been developed that can be used in addition to or as an enhancement to established techniques. The analysis of body fluids using RNA as a target, which complements classical presumptive and confirmatory tests, has yielded promising results. The RNA-based approach enhances the possibility of classifying biological material found at crime scenes more precisely than as "blood, saliva, or semen"; for example using RNA-based approach, it is possible to differentiate between menstrual and circulatory blood, which can be critical in sexual assault cases.

Advances in our understanding of the human genome have made a range of markers, in particular single nucleotide polymorphisms (SNPs) and insertion-deletion polymorphisms (INDELs), easier to identify. Both these categories of markers have an advantage over STRs when analyzing poor-quality DNA as they can be amplified on shorter amplicons, thereby increasing the chance of successful analysis, especially from degraded DNA. While their use is not currently widespread in the forensic community, both SNPs and INDELs can be used as a supplementary tool along with STRs when the power of discrimination needs to be increased; panels of markers have been selected and validated to address this need.

Alternative panels of SNPs and INDELs have also been selected that can provide an estimation of ancestry, rather than for identification, and have the critical feature that their frequencies of occurrence differ between different major population groups, allowing probabilistic estimations of the geographical origin or biological material, or genetic ancestry, to be made. The information gained from this type of analysis is not nearly as powerful as from direct matching to a reference profile but can provide investigators with a lead, allowing them to focus their investigations on fewer people. A different approach to providing investigators with some information on an unknown sample is to try and build up a phenotypic portrait of the individual that the biological material came from. SNPs have been identified that contribute to externally visible traits, such as eye and hair color, and can potentially be applied to casework, but only in certain populations, where eye and hair color vary significantly.

At the time of writing, only one commercial kit is available for the analysis of autosomal INDELs and none for SNPs (using capillary electrophoresis). This presents a barrier to the widespread adoption of these markers. Next Generation Sequencing (NGS) technology could provide a platform that facilitates the use of alternative markers in more laboratories. NGS has been used for genome analysis for several years but has not been particularly suited to forensic analysis as it has been expensive and requires relatively high amounts of good-quality DNA. However, with the development of cheaper and more robust platforms, it

offers the potential to type multiple SNPs, INDELs, and STRs in a single reaction, providing a vast amount of information for every sample typed. The first commercial kits designed specifically for forensic investigations have been released by Illumina and Life Technologies for use on their NGS platforms and are being evaluated in several laboratories around the world.

Developments in Y chromosome STR analysis and mitochondrial analysis (mtDNA) continue to be made, for example, the identification of rapidly mutating Y STRs that allow for better discrimination between males from the same paternal lineage. NGS-based analysis offers a route for sequencing more, or all, of the mitochondrial genome, which also increases it potential for discrimination.

This book presents a series of protocols. It is envisaged that some of these will be already routine in some laboratories, but may be of use to other laboratories that have not yet implemented or possibly considered the method. Even when the method is already familiar, comparison with practice in another laboratory is valuable. Others methods are not as widely used and as such the methodology presented, and in particular the notes that add extra detail to that given in published papers, can hopefully assist with the execution of the different techniques.

The book is arranged into a series of related chapters. The start point is the collection and storage of biological material: as DNA profiling gets more sensitive the collection, storage, and processing need to reduce the possibility of inadvertently contaminating the sample or contributing to its deterioration. Following on from this Chapters 2–3 examine two different aspects of RNA analysis for body fluid identification. Methods for the storage of biological material and DNA extraction are covered in Chapters 4–7, the focus of Chapters 6 and 7 is the extraction of DNA from hard tissues. Chapters 8–10 present methods for monitoring the quality of DNA extracts and also further steps to aid with the purification of DNA, making it amenable to downstream analysis. Chapters 11–16 detail methods for the typing on nonstandard markers, including INDELs, Y chromosome STRs, and mitochondrial DNA. Chapters 17–19 deal with phenotype and ancestry, including detailed procedures for data analysis and evaluation. Finally, Chapter 20 looks at the application of DNA typing to the identification of non-human material to species level. Detailed NGS protocols have not been included in this book as it was felt that the protocols were not sufficiently standardized, with many laboratories assessing different parameters involved in the generation of data. However, many of the protocols presented can feed into an NGS workflow.

I am indebted to all of the authors that have shared their experience and provided the manuscripts that make up this book. I would like to thank Professor John Walker, the series editor, for providing prompt guidance and advice throughout the process of compiling this work.

Preston, Lancashire, UK *William Goodwin*

Contents

Contributors

Wafa Altayari • *Abu Dhabi Police, Forensic Evidence Department, Abu Dhabi, United Arab Emirates*

J. Alvarez-Dios • *Faculty of Mathematics, University of Santiago de Compostela, Galicia, Spain*

Jack Ballantyne • *National Center for Forensic Science, Orlando, FL, USA; Department of Chemistry, University of Central Florida, Orlando, FL, USA*

Kaye N. Ballantyne • *Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, Melbourne, Victoria, Australia*

David Ballard • *Faculty of Life Sciences and Medicine, King's College London, London, UK*

Majid Bashir • *The Forensic Laboratory Department, Ministry of Interior, Doha, Qatar*

B. Bekaert • *Department of Imaging & Pathology, Forensic Biomedical Sciences, KU Leuven —University of Leuven, Leuven, Belgium; Department of Forensic Medicine, Laboratory of Forensic Genetics and Molecular Archaeology, University Hospitals Leuven, KU Leuven — University of Leuven, Leuven, Belgium*

Ángel Carracedo • *Forensic Genetics Unit, Luis Concheiro Institute of Forensic Sciences, Genomic Medicine Group, University of Santiago de Compostela, Galicia, Spain; Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah, Saudi Arabia*

R. Decorte • *Department of Imaging & Pathology, Forensic Biomedical Sciences, KU Leuven — University of Leuven, Leuven, Belgium; Department of Forensic Medicine, Laboratory of Forensic Genetics and Molecular Archaeology, University Hospitals Leuven, KU Leuven — University of Leuven, Leuven, Belgium*

Kenneth Drees • *Department of Biology, University of North Dakota, Grand Forks, ND, USA*

Suni M. Edson • *Armed Forces DNA Identification Laboratory (AFDIL), American Registry of Pathology (ARP) Contractors supporting the Armed Forces medical Examiner System (AFMES), Dover AFB, DE, USA*

R. Ellerington • *Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK*

Manuel Fondevila • *Forensic Genetics Unit, Luis Concheiro Institute of Forensic Sciences, Genomic Medicine Group, University of Santiago de Compostela, Galicia, Spain*

A. Gomez-Tato • *Faculty of Mathematics, University of Santiago de Compostela, Galicia, Spain*

William Goodwin • *School of Forensic and Investigative Science, University of Central Lancashire, Preston, UK*

Cordula Haas • *Institute of Legal Medicine, University of Zurich, Zurich, Switzerland*

Sibte Hadi • *School of Forensic and Investigative Sciences,University of Center Lancashire, Preston, UK*

Erin K. Hanson • *National Center for Forensic Science, Orlando, FL, USA*

Nur Haliza bt Hassan • *Department of Chemistry Malaysia, Jalan Sultan, Selangor, Malaysia*

William R. Hudlow • *California department of Justice, Jan Bashinski DNA Laboratory, Richmond, CA, USA; California Criminalistics Institute, Rancho Cordova, CA, USA*
Joanna Jakubowska • *Department of Forensic Medicine, Laboratory of Forensic Biology and Genetics, Medical University of Gdansk, Gdańsk, Poland*
Manfred Kayser • *Department of Forensic Molecular Biology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands*
Katelyn Kjelland • *Department of Biology, University of North Dakota, Grand Forks, ND, USA*
Maria Victoria Lareu • *Forensic Genetics Unit, Luis Concheiro Institute of Forensic Sciences, Genomic Medicine Group, University of Santiago de Compostela, Galicia, Spain*
James Chun-I Lee • *Department of Forensic Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan, ROC*
Adrian Linacre • *Department of Biological Sciences, Flinders University, Adelaide, South Australia*
Agnieszka Maciejewska • *Department of Forensic Medicine, Laboratory of Forensic Biology and Genetics, Medical University of Gdansk, Gdańsk, Poland*
Mathew J. Malek • *Department of Biology, University of North Dakota, Grand Forks, ND, USA*
Timothy P. McMahon • *Armed Forces DNA Identification Laboratory (AFDIL), American Registry of Pathology (ARP), Contractors supporting the Armed Forces Medical Examiner System (AFMES), DE, USA*
Dennis McNevin • *National Centre for Forensic Studies, Faculty of Education, Science, Technology and Mathematics, University of Canberra, Canberra, Australia*
Roland A.H. van Oorschot • *Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, Melbourne, Victoria, Australia*
Igor V. Ovchinnikov • *Department of Biology, University of North Dakota, Grand Forks, ND, USA*
Irena Zupanič Pajnič • *Laboratory of Molecular Genetics, Institute of Forensic Medicine, Ljubljana, Slovenia*
Ryszard Pawlowski • *Department of Forensic Medicine, Laboratory of Forensic Biology and Genetics, Medical University of Gdansk, Gdańsk, Poland*
Chris Phillips • *Forensic Genetics Unit, Luis Concheiro Institute of Forensic Sciences, Genomic Medicine Group, University of Santiago de Compostela, Galicia, Spain*
Amy D. Roeder • *Cellmark, Blacklands Way, Abingdon, Oxfordshire, UK*
Carla Santos • *Forensic Genetics Unit, Luis Concheiro Institute of Forensic Sciences, genomic Medicine Group, University of Santiago de Compostela, Galicia, Spain*
L. Van den Abbeele • *Faculty of Medicine and Pharmacy, Free University of Brussels, Brussels, Belgium*
Timothy J. Verdon • *Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, Melbourne, Victoria, Australia*
Susan Walsh • *Department of Biology, Indiana University-Purdue University Indianapolis (IUPUI), Indianapolis, IN, USA*
Nathalie Zahra • *School of Forensic and Investigative Science, University of Central Lancashire, Preston, UK*

Chapter 1

Collection of Samples for DNA Analysis

Roland A.H. van Oorschot, Timothy J. Verdon, and Kaye N. Ballantyne

Abstract

Effective sampling of biological material is critical to the ability to acquire DNA profiles of probative value. The main methods of collection are swabbing, tapelifting, or direct excision. This chapter describes the key aspects to consider when applying these methods, in addition to suggested procedures for swabbing and tapelifting. Important issues to be considered, such as exhibit triaging, pre-examination preparation, contamination risk reduction, sample localization, sample identification, and sample prioritization as well as aspects of record keeping, packaging, and storage, are also raised.

Key words DNA collection, Sample targeting, Contamination risk, Swabbing, Tapelifting, Forensic

1 Introduction

The collection of biological material for DNA analysis is more than just about dabbing a swab at an obvious stain, putting it in a tube, and submitting it for DNA analysis. Identifying what needs to be collected and the most effective method of collection are crucial aspects of the DNA analysis process. Incorrect or incomplete targeting and collection of relevant material will limit the probability of acquiring a useful DNA profile, the quality and probative value of the profile generated, and ultimately its contribution to assisting the justice system.

The issues surrounding collection of DNA evidence are far greater than just methods for the identification and sampling of biological material. Analysts must be cognizant of the timing of DNA examinations relative to other forensic disciplines' examinations in terms of impacts on each other, the risk of contamination prior to or during examination or sampling, the need for presumptive and confirmatory tests prior to sampling, as well as record keeping, packaging, and storage.

This chapter will focus on the main collection methods of biological material for DNA analysis including whole or partial substrate collection, swabbing, and tapelifting (*see* **Note 1**).

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_1,

However, information regarding the additional considerations in forensic evidence identification and recovery is also provided, to allow readers a full comprehension of the complex nature of the collection task. It is important to note that comparatively little empirical research has been performed in the collection area, and as such many common practices are based on assumptions, anecdotal evidence, or practitioner's experience, rather than on confirmed, controlled, and tested data. Wherever possible within this chapter, the reader will be directed to relevant references for such data.

2 Exhibit Triaging, Pre-examination Preparation, and Sample Targeting

2.1 Multidisciplinary Evidence Recovery

For many scenes or exhibits, DNA is not the only evidence available or desired for forensic investigation. However, the examination and collection of biological material may impact on the recovery of other evidence. It may remove other forms of trace evidence such as fibers, glass, or plant material; obliterate indentations or writing prior to document analysis; damage electronic equipment; remove or obscure fingerprints and toolmarks, and complicate damage and blood pattern analysis. Therefore, appropriate communication with the informant and other relevant forensic discipline specialists should be performed prior to any examination. The relative order of evidence collection may be highly case specific based on the desired analyses. In general, adopting a trace evidence first strategy is considered beneficial to prevent contamination or loss of minute or easily transferred evidence such as trace DNA, gunshot residue, or fibers through excessive handling or manipulation.

Whilst considering the relative order of examinations, thought must be given not only to the impact of DNA collection on other examinations but also to the impact of other examinations on potential DNA evidence. For example, the use of fingerprinting brushes and powder to visualize a fingerprint may reduce the opportunity to obtain a DNA profile from a print (regardless of any print providing probative value or not) due to removal of sample [1] or potential contamination of the target area [1, 2]. Some fingerprinting methods may even inhibit the DNA profiling process [3–7]. The use of electrostatic detection apparatus (ESDA) for indentations or stubs for gunshot residue collection may remove significant amounts of DNA present on the item.

2.2 Contamination Risk Reduction

Minimizing contamination risk during examination and collection is vital for preserving the integrity of DNA evidence. Next generation profiling kits such as PowerPlex ESI/X and GlobalFiler provide such exceptional sensitivity that even minute background traces

or contaminants can be detected. Therefore, strict procedures must be in place to prevent the loss or obfuscation of evidence.

The same level of contamination risk reduction procedures should be adopted when undertaking other pre-DNA collection examinations of an exhibit. One could argue that they should also be adopted when undertaking any post-DNA examinations of an exhibit to preserve the integrity of the exhibit in case of any desired further/re-examinations.

If undertaking the sample collection in the field, secure the area around the exhibit and restrict access as soon as possible. If within a laboratory environment, areas for sample examination should be isolated or semi-isolated with limited traffic flow within the vicinity, and have near neutral air flow in the area of the examination bench.

Prior to entering the examination area, whether at a scene or within a laboratory, full protective equipment should be put on by any personnel approaching the scene or bench. These include a lab coat/gown, mask covering nose and mouth, hairnet, and gloves. These clothing items protect the exhibit from biological material (e.g., saliva from talking/coughing, loose hairs, shed skin cells on clothing, and oils/loose DNA on skin) being transferred to an exhibit. The gloves, which are the protective equipment most likely to come into contact with an exhibit, must be DNA free. Care should be taken when removing the gloves from their packaging to prevent contact with bare hands/skin (especially the finger and palm areas of the new glove being removed) or potentially contaminated surfaces such as the exterior of the glove box. The practice of double gloving where a second glove is worn over the first can further assist with minimizing this contamination risk.

Attention should also be paid to the order in which protective clothing is applied—the hairnet and face mask should be put on first (to prevent hair/saliva etc. falling onto the lab gown), then the gown itself, then gloves, ensuring that the gloves cover the wrist and forearm, overlapping the sleeve of the gown.

Within the laboratory, the examination area should be supplied with only the tools and equipment required for the specific task to ensure an uncluttered environment with minimized risk of contamination. All equipment and surfaces should be thoroughly cleaned before and after each use. The classification of items as high, medium, or low risk can aid in raising awareness of contamination probabilities, and ensure the appropriate level of care is taken for each item. High risk items are those that may come into direct contact with an exhibit (e.g., forceps, scissors, gloves); medium risk may function as a one-step vector between a non-exhibit surface and an exhibit (e.g., glove box, pen, magnifying lamp, water dropper) [8, 9]. Lower contamination risk surfaces require multiple transfer steps for DNA to be transferred to an exhibit, commonly involving a medium or high risk vector. Whilst

all surfaces should be cleaned regularly, focus should be on high and medium risk items with cleaning before and after each use. Cleaning agents should be of sufficient strength to ensure eradication of surface DNA, but able to be removed entirely from the item to prevent transfer and destruction of target DNA [10]. Commonly used agents include sodium hypochlorite (0.5–10 % v/v solutions [11, 12]), UV [13], and ethylene oxide [14], although numerous other commercial solutions are available. The use of disposable, certified DNA-free items such as scissors, forceps, and racks is recommended—no cleaning method is 100 % effective, and so care should be exercised. Separate areas should be designated for the exhibit and cleaned tools, away from potentially contaminated items such as exhibit packaging, paperwork, and biohazard waste/rubbish. The positioning of such items should also be such that there will not be any traffic over an exhibit when they are to be utilized during an examination.

Gloves should be changed regularly throughout the examination, particularly after contact with exhibit (to prevent transferring DNA from the exhibit to laboratory surfaces), and after contact with potentially unclean medium or low risk items such as laboratory furniture, paperwork, and packaging. Likewise, if possible separate tools should be used on separate stains—using forceps or scissors on a trace sample immediately after a blood stain may create an artificial mixture of both stains [8, 15].

When handling an exhibit do so as little as possible and try to avoid areas likely to be targeted for sampling.

To assist in determining the possible source of contamination, in situations where this is deemed necessary, it is imperative to (a) record which benches, tools, and equipment were used, by who and when, per case/exhibit; and (b) have a current and comprehensive staff elimination database of DNA profiles, of all typing systems in use within a laboratory, of all relevant individuals.

2.3 Sample Targeting and Ordering

The area to be sampled needs to be clearly defined. This will be dependent on the question(s) being asked from the investigative perspective. Awareness of the hypotheses to be tested in light of the context of the case can assist in identifying relevant (or irrelevant) stains and traces. Visualization or enhancement techniques such as alternative light sources [16–18] and/or near infrared cameras [19] can assist in locating possible biological material. Prior to sampling for DNA analysis presumptive or confirmatory tests can be used to screen non-relevant stains and prioritize sampling areas. When possible, a test or method that does not come into contact with a potential area to be sampled should be employed, to minimize loss of sample and risk of contaminating the sample (*see* **Note 2**).

The ordering of sample collection should be based on (a) samples most likely to provide relevant information (taking into account biological sample type, quantity of sample, age of sample,

sample location), (b) the nature of the sample—trace samples should be collected first as these are most prone to/affected by contamination, (c) the likely yield of DNA from specific volumes or sizes of stains of biological materials, and the relevant collection or extraction efficiency for the substrate and collection method employed (*see* **Note 3**).

It may be necessary to section the stain or exhibit and collect separate samples to prevent sample mixing or spreading. An example of incorrect sample mixing would be swabbing the handle and blade of a knife together—in many instances in doing so victim and perpetrator DNA may be mixed, confusing the correct attribution of the location of each individual on the item. Spreading of a trace sample beyond the target area may also reduce recovery yield and thus ability to acquire a profile of sufficient probative value.

Additionally, consideration must be given to jurisdiction or laboratory specific requirements for number of samples allowed per exhibit or case type, the need for additional testing by independent laboratories, or the potential need for re-examination subsequent to further investigation.

3 Sampling

3.1 Choice of Sampling Method

The choice of sampling method or collection device is highly dependent on the type and size of exhibit, the surface which needs to be sampled, and the location of the target deposit. Choices include:

1. Excision of portion of exhibit or use of whole exhibit for direct extraction, for fabrics or small items such as cartridge casings [20].
2. Swabbing: mainly used to collect biological material from nonporous surfaces, e.g., knife blades and handles, bottles, mugs, watches, ceramic tiles. *See* Subheading 3.2.
3. Tapelifting: mainly used to collect biological material from porous surfaces (mainly fabric substrates), e.g., clothing, bedding. *See* Subheading 3.3.

As some collection devices are more amenable to specific extraction methods than others due to substrate type or size limitations, an understanding and consideration of the subsequent DNA extraction methods available and preferred within the laboratory is important.

The invasiveness of the sampling method and the preservation of the exhibit or item should be considered relative to the likely quantity and quality of DNA that will be retrieved. If, for example, swabbing a blood stain on clothing will provide sufficient DNA to obtain a full profile, excising a section of the stain could be considered unnecessary damage to the exhibit for the owner, and could complicate further examination at a later date.

3.2 Swabbing

There are numerous choices for swabs available for the collection of biological material. The use of different swabs for different materials and substrates can be beneficial in different circumstances [21], although data are still scarce for many situations or combinations. *See* Table 1 for suggestions of swab selection for sampling different forms of biological evidence from different substrate types. Many laboratories use a single type of swab, most commonly cotton [22], for all purposes. To collect the maximum amount of material via swabbing, the following procedure is recommended:

1. Thoroughly dampen the swab with nuclease-free water (*see* **Note 4**), but do not oversaturate to the point of dripping. The vessel of water or dropper used should not contact the swab to prevent contamination.
2. Hold the swab stick such that you can apply some pressure of the swab onto the substrate. Your fingers should however not be placed too close to the swab (i.e., avoid contact with the swab bud) and the swab stick must be held in such a way that ensures that the fingers, palm, and/or wrist do not touch the substrate from which you are collecting the DNA.
3. Traverse the total target area multiple times applying a medium level of pressure. Whilst doing this keep the swab at an angle relative to the substrate, such that a large area of the swab is contacting the substrate, and continually rotate the swab.
4. Ensure all the moisture is absorbed by the swab.
5. Empirical evidence has shown that a single swab is insufficient to collect all available DNA on a substrate [23–26]. **Steps 1–3** should thus be repeated with a second swab ensuring that no moisture is left on the substrate. The second swab should be co-extracted with the first (in the same tube where possible) to

Table 1
Suggested swab types which have been shown to effectively sample a variety of forms of biological evidence from different substrates [21]

Evidence	Touch				Biological fluid stain			
Porosity	**Porous**		**Nonporous**		**Porous**		**Nonporous**	
Texture	Rough	Smooth	Rough	Smooth	Rough	Smooth	Rough	Smooth
Swab type	Cotton wound	Foam sleeve	Cotton wound	Polyester wound	Rayon wound	Foam sleeve	Polyester wound	Cotton wound

Porosity and texture relate to the ability of the substrate to allow liquid to seep through and the degree of roughness of a substrate, respectively. Wound swabs comprise long fibers (generally cotton, polyester, or rayon) tightly wrapped around the swab stick; sleeve swabs consist of a sleeve of foam or woven fibers affixed over the end of the swab stick. For relative performances of different swab types on different substrates *see* Ref. [21]

retrieve as much of the collected DNA as possible. Whilst some laboratories recommend the second swab to always be left dry rather than dampened, it is unlikely to collect much additional material from a dry substrate. So the only occasions where this could be admissible is if for some reason the substrate is still very moist after completing the first swabbing (*see* **Note 5**).

6. If the surface area to be swabbed is very large, to the point where a swab appears to become dry well before the whole area has been traversed, additional swabs should be used. This should continue until the whole target area has been swabbed in an equal manner. All swabs should be extracted together or if this is not practical or inefficient in separate aliquots and the extracted DNA subsequently pooled.

3.3 Tapelifting

1. Choose a tape which is either DNA-free or at least able to be cleaned before use. It has been suggested that a stickier tape will collect more DNA of a higher quality than a less adhesive tape [21, 27]. Also ensure that DNA can be extracted and analyzed effectively from the tape you have chosen with the methods used in your laboratory.
2. If using pre-cut, commercially available tapelifters, proceed to **Step 4**. If using a roll of tape, choose a length which is appropriate to the sampling area but not too much to make extraction difficult. A length of approximately 5 cm (2″) is a suggested maximum length. Cut a length of tape longer than the chosen size by 5–10 cm (2–4″).
3. Using the additional length, fold both sticky ends of the tape back to create two non-adhesive handles (each approximately a quarter of the additional length cut) on either side of the section of the tape to be used for sampling. The tapelifter should now have an adhesive surface in the middle, bordered by two non-adhesive handles on either side.
4. Apply the adhesive portion of the tape to the substrate, pressing firmly along the back of the tape. Lift the tape off and apply to an adjacent region of the area to be sampled. Continue to apply to adjacent areas within the target area.
5. Once the whole target area has been sampled, repeat sampling until the entire target area has been sampled several times. It has been observed that sampling an area twice the size of the tapelifter approximately eight times provides a relatively good rate of recovery [27] (*see* **Note 6**).
6. If at any stage prior to completing the collection the tape appears to no longer adhere to the substrate, continue with a fresh tapelifter.
7. Remove the handles using a DNA-free scalpel or scissors, and extract DNA directly from the adhesive portion of the tape

which contains the biological material. If more than one tape-lift is collected, they should be extracted together or if this is not practical or inefficient in separate aliquots and the extracted DNA subsequently pooled.

4 Record Keeping, Packaging, and Storage

4.1 Record Keeping

Information must be recorded regarding what was collected (the assumed biological material), from which exhibit, from which location on the exhibit, using what type of collection device, when, and by whom. This information can be collected by different means including combinations of handwritten or electronically typed notes and drawings, photos, video, and/or voice recordings. Many jurisdictions have specific document controlled forms and associated procedures available to assist recording of relevant information. Apart from its value within the particular case investigation for which the sample was collected, collating collected information of this type for all samples collected within a laboratory or jurisdiction can assist in regular benchmarking assessments that can help identify and drive improvements in success rates of samples collected (acquisition of data of probative value), quality of the samples, efficiencies and training, as well as identify research directions [28].

Post-examination and sampling, the collection device and/or the packaging it is to be stored in needs to be labeled with the appropriate identification details (exhibit/sample number or barcode) that can be cross-referenced to the relevant details recorded.

4.2 Packaging and Storage

Any samples which are still moist after collection should be dried as soon as possible to prevent fungal and bacterial growth that will harm the future analyses of the collected target DNA. Alternatively they should be frozen immediately.

In respect to swabs, some manufactures provide pre-perforated containers, or containers with desiccant, to assist drying. However, many do not, so it is essential that the containers are modified to allow air drying of the swab. This could be done by cutting a hole in the container in the vicinity of the swab head. A DNA-free implement (scissors, scalpel) should be used for this task to prevent contamination. Any modifications to the container should still ensure that the swab head is not able to contact any surface other than the internal surface of the container. To prevent swabs in semi-open containers becoming contaminated they should be placed in individual paper envelopes/bags. They should not be placed in plastic bags as they may prevent drying. Swabbings of the same target area of an exhibit can be placed within the same envelope but those of other target areas must be placed in separate envelopes. All envelopes are to be appropriately identified.

In most laboratories envelopes with swabs can be stored at room temperature; however, in areas of high humidity it is advised that storage rooms with lower humidity are used. Any labels to be used must not be able to come loose, and the ink used must not smudge or fade if it becomes wet or exposed to freezing conditions.

5 Notes

1. Other less utilized collection methods exist that can be useful in specific situations such as vacuuming which may be a method of choice when wanting to collect trace samples from large fabric surfaces such as car seats or bedding [29] and laser microdissection which may be used to isolate a small number of relevant target cells away from the bulk of other cells whose DNA would swamp that of the target cells if co-extracted [30–32].
2. Most jurisdictions currently still perform pre-sampling presumptive tests such as Hemastix (Siemens Healthcare Diagnostics), seminal acid phosphatase test, and Phadebas (Magle Life Sciences) for blood, semen, and saliva respectively, but many laboratories in addition utilize post-sampling confirmatory tests such as ABAcard® Hematrace® (Abacus Diagnostics) for blood, RSID™-Saliva (Independent Forensics) for saliva, and/or ABAcard® p30 (Abacus Diagnostics) or RSID™-Semen (Independent Forensics) for semen. Some of the presumptive and confirmatory tests require the use of valuable sample and separate tests for different biological materials. Some laboratories are starting to consider alternative methods using mRNA for post-sampling biological sample identification. Such methods are able to not only deduce the type of biological material for specific types of biological fluids and tissues for which routine immunological/enzymatic/chemical presumptive or confirmatory tests are available but also several other relevant biological fluid and tissues for which routine tests are not available, such as for menstrual blood and skin [33–35]. Extraction of mRNA from the biological sample can be done simultaneously with DNA extraction and the extracted mRNA typed prior to DNA profiling, or if needed after DNA profiling [36, 37].
3. The amount of DNA retrievable from a given volume, or size of stain, varies depending on the biological source. Per volume of fluid or associated stain one can expect to obtain most from semen (~150–300 μg/ml), followed by blood (~20–40 μg/ml, 250–500 ng/cm^2), then saliva (~1–10 μg/ml) with the amount of DNA left by touch on a hand-sized area usually

providing far less (<0.1–169 ng) than all of the above [38–40]. Whilst touch DNA is often relatively low it should not by definition be considered a trace sample. "Trace sample" or "trace DNA" refers to any sample which may fall below the recommended threshold at any stage of the process—detection, collection, extraction, amplification, and interpretation [41].

4. Alternative wetting agents may be used in place of water, and may provide enhanced recovery in some circumstances [42, 43]. Further research may reveal alternative wetting agents that provide the sought-after enhanced recovery. Any such alternative should be implemented into standard operating procedures.
5. Depending on the type of swab and the stick it is attached to, it is usually preferable for the swab to be removed from the stick prior to placing it in the extraction tube. Also, depending on the type of swab, especially those made of tightly wound material such as the commonly used cotton swabs and when the swab has been allowed to dry prior to commencement of the DNA extraction process, teasing loose the swab head prior to placement in the extraction tube may assist retrieval of DNA during the extraction process.
6. The most effective number of times to sample a substrate will depend on the substrate type, size, and the dimensions of the tapelifter. Applying tape too many times has been shown to diminish the quantity of DNA in the extract [27], and tapelifting will generally be less successful from fabrics with a large number of detachable fibers, such as flannelette. Taping "until no longer sticky," often quoted as an endpoint for tapelifting, may lead to under- or over-tapelifting and should be discouraged.

Disclaimer

The views presented in this chapter are solely those of the authors and do not necessarily represent views or policies of their respective organizations.

References

1. van Oorschot RAH, Treadwell S, Beaurepaire J, Holding NL, Mitchell RJ (2005) Beware of the possibility of fingerprinting techniques transferring DNA. J Forensic Sci 50:1417–1422
2. Gibb C, Gutowski SJ, van Oorschot RAH (2012) Assessment of the possibility of DNA accumulation and transfer in a superglue chamber. J Forensic Ident 62:409–424
3. van Hoofstat DE, Deforce DL, Hubert De Pauw IP, van den Eeckhout EG (1999) DNA typing of fingerprints using capillary electrophoresis: effect of dactyloscopic powders. Electrophoresis 20:2870–2876

4. Stein C, Kyeck SH, Henssge C (1996) DNA typing of fingerprint reagent treated biological stains. J Forensic Sci 41:1012–1017
5. Roux C, Gill K, Sutton J, Lennard C (1999) A further study to investigate the effect of fingerprint enhancement techniques on the DNA analysis of bloodstains. J Forensic Ident 49:357–376
6. Frégeau CJ, Germain O, Fourney RM (2000) Fingerprint enhancement revisited and the effects of blood enhancement chemicals on subsequent profiler plus(TM) fluorescent short tandem repeat DNA analysis of fresh and aged bloody fingerprints. J Forensic Sci 45:354–380
7. Raymond JJ, Roux C, Du Pasquier E, Sutton J, Lennard C (2004) The effect of common fingerprint detection techniques on the DNA typing of fingerprints deposited on different surfaces. J Forensic Ident 54:22–44
8. Poy AL, van Oorschot RAH (2006) Trace DNA presence, origin, and transfer within a forensic biology laboratory and its potential effect on casework. J Forensic Ident 56: 558–576
9. Ballantyne KN, Poy AL, van Oorschot RAH (2013) Environmental DNA monitoring: beware of the transition to more sensitive typing methodologies. Aust J Forensic Sci 45:323–340
10. Bright JA, Cockerton S, Harbison S, Russell A, Samson O, Stevenson K (2011) The effect of cleaning agents on the ability to obtain DNA profiles using the identifiler™ and PowerPlex® Y multiplex kits. J Forensic Sci 56:181–185
11. Harris KA, Thacker CR, Ballard D, Court DS (2006) The effect of cleaning agents on the DNA analysis of blood stains deposited on different substrates. Int Congr Ser 1288:589–591
12. Vandewoestyne M, van Hoofstat D, De Groote S, van Thuyne N, Haerinck S, van Nieuwerburgh F et al (2011) Sources of DNA contamination and decontamination procedures in the forensic laboratory. J Forensic Res S2:001. doi:10.4172/2157-7145.S2-001
13. Tamariz J, Voynarovska K, Prinz M, Caragine T (2006) The application of ultraviolet irradiation to exogenous sources of DNA in plasticware and water for the amplification of low copy number DNA. J Forensic Sci 51:790–794
14. Shaw K, Sesardic I, Bristol N, Ames C, Dagnall K, Ellis C et al (2008) Comparison of the effects of sterilisation techniques on subsequent DNA profiling. Int J Legal Med 122:29–33
15. Szkuta B, Harvey ML, Ballantyne KN, van Oorschot RAH (2013) The potential transfer of trace DNA via high risk vectors during exhibit examination. Forensic Sci Int Genet Suppl Ser 4:e55–e56
16. Lennard C, Stoilovic M (2004) Application of forensic light sources at the crime scene. In: Horswell J (ed) The practice of crime scene investigation. CRC Press, Boca Raton, FL, pp 97–124
17. Vandenberg N, van Oorschot RAH (2006) The use of polilight((R)) in the detection of seminal fluid, saliva, and bloodstains and comparison with conventional chemical-based screening tests. J Forensic Sci 51:361–370
18. Ben Yosef N, Almog J, Frank A, Springer E, Cantu AA (1998) Short UV luminescence for forensic applications: design of a real-time observation system for detection of latent fingerprints and body fluids. J Forensic Sci 43:299–304
19. Albanese J, Montes R (2011) Latent evidence detection using a combination of near infrared and high dynamic range photography: an example using bloodstains. J Forensic Sci 56:1601–1603
20. Dieltjes P, Mieremet R, Zuniga S, Kraaijenbrink T, Pijpe J, De Knijff P (2011) A sensitive method to extract DNA from biological traces present on ammunition for the purpose of genetic profiling. Int J Legal Med 125:597–602
21. Verdon TJ, Mitchell RJ, van Oorschot RA (2014) Swabs as DNA collection devices for sampling different biological materials from different substrates. J Forensic Sci 59(4):1080–1089
22. Raymond JJ, van Oorschot RAH, Walsh SJ, Roux C (2008) Trace DNA analysis: do you know what your neighbour is doing? A multijurisdictional survey. Forensic Sci Int Genet 2:19–28
23. van Oorschot RAH, Weston RK, Jones MK (1998) In: Retrieval of DNA from touched objects. In: Proceedings of the 14th International Symposium of Forensic Sciences of the Australian and New Zealand Forensic Science Society, Adelaide
24. van Oorschot RAH, Szepietowska I, Scott DL, Weston RKM, Jones J (1999) In: Retrieval of genetic profiles from touched objects. In: Proceedings of the First International Conference on Forensic Human Identification in the Millennium, London. Forensic Science Service (UK)
25. Sweet D, Lorente M, Lorente JA, Valenzuela A, Villanueva E (1997) An improved method

to recover saliva from human skin: the double swab technique. J Forensic Sci 42:320–322

26. Pang BCM, Cheung BKK (2007) Double swab technique for collecting touched evidence. Leg Med 9:181–184
27. Verdon TJ, Mitchell RJ, van Oorschot RAH (2014) Evaluation of tapelifting as a collection method for touch DNA. Forensic Sci Int Genet 8:179–186
28. van Oorschot RAH (2012) Assessing DNA profiling success rates: need for more and better collection of relevant data. Forensic Sci Pol Manag 3:37–41
29. Garrett AD, Patlak DJ, Gunn LE, Brodeur AN, Grgicak CM (2014) Exploring the potential of a wet-vacuum collection system for DNA recovery. J Forensic Ident 64:429–448
30. Anoruo B, van Oorschot RAH, Mitchell RJ, Howells D (2007) Isolating cells from non-sperm cellular mixtures using the PALM® microlaser micro dissection system. Forensic Sci Int 173:93–96
31. Anslinger K, Bayer B, Mack B, Eisenmenger W (2007) Sex-specific fluorescent labelling of cells for laser microdissection and DNA profiling. Int J Legal Med 121:54–56
32. Vandewoestyne M, van Hoofstat D, van Nieuwerburgh F, Deforce D (2009) Suspension fluorescence in situ hybridization (S-FISH) combined with automatic detection and laser microdissection for STR profiling of male cells in male/female mixtures. Int J Legal Med 123:441–447
33. Haas C, Klesser B, Maake C, Bär W, Kratzer A (2009) mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. Forensic Sci Int Genet 3:80–88
34. Fleming RI, Harbison S (2010) The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids. Forensic Sci Int Genet 4:244–256
35. Lindenbergh A, De Pagter M, Ramdayal G, Visser M, Zubakov D, Kayser M et al (2012) A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces. Forensic Sci Int Genet 6:565–577
36. Bauer M, Patzelt D (2003) A method for simultaneous RNA and DNA isolation from dried blood and semen stains. Forensic Sci Int 136:76–78
37. Bowden A, Fleming R, Harbison S (2011) A method for DNA and RNA co-extraction for use on forensic samples using the promega DNA IQ™ system. Forensic Sci Int Genet 5:64–68
38. Lee HC, Ladd C (2001) Preservation and collection of biological evidence. Croat Med J 42:225–228
39. van Oorschot RAH, Jones MK (1997) DNA fingerprints from fingerprints [6]. Nature 387:767
40. Daly DJ, Murphy C, McDermott SD (2012) The transfer of touch DNA from hands to glass, fabric and wood. Forensic Sci Int Genet 6:41–46
41. van Oorschot RAH, Ballantyne KN, Mitchell RJ (2010) Forensic trace DNA: a review. Investig Genet 1:14
42. Prinz M, Schiffner L, Sebestyen JA, Bajda E, Tamariz J, Shaler RC, et al (2006) Maximization of STR DNA typing success for touched objects. Int Congr Ser 1228:651–653
43. Collopy C (2008) Mini-popule developed to maximize DNA recovery for robotic forensic analysis. Forensic Mag 01/02/2008

Chapter 2

Body Fluid Identification Using mRNA Profiling

Amy D. Roeder and Cordula Haas

Abstract

RNA analysis is a valuable tool for the identification of the forensically relevant body fluids, saliva, blood, menstrual blood, cervicovaginal fluid, and semen. Multiple human mRNA and bacterial RNA markers have been identified for each of these body fluids. RNA and DNA can be coextracted from the same portion of a sample and RNA markers for different body fluids can be multiplexed in a single PCR, thereby maximizing the number of analyses that can be performed with limited sample material.

Key words Forensic, Body fluid identification, mRNA profiling

1 Introduction

In some instances, the composition of a biological sample is critical for the interpretation of forensic evidence. Although DNA analysis is widely used to determine who may have left biological evidence, at present DNA-based methodology cannot be used to identify all of the forensically relevant body fluids. Currently, immunological or enzymatic assays are widely used to identify saliva, blood, and semen. There are also commercially available assays for the identification of urine and menstrual blood (MB) and published histological methods for the identification of vaginal cells. However, to varying degrees, all of these tests are known to yield false positive results. Furthermore, the assay(s) for each body fluid are performed separately, so additional sample material is consumed with each test performed.

Messenger RNA (mRNA) profiling relies on the differential expression of mRNAs in different tissues to identify the presence of a body fluid/tissue. For forensic purposes, mRNA markers have been described for semen, blood, MB, cervicovaginal fluid (CVF), saliva, and skin [1–11]. Unlike the enzymatic or immunologic assays, mRNA markers for multiple body fluids/tissues can be multiplexed together in a single PCR-based assay. RNA and DNA can be coextracted from the same portion of the sample. mRNA

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_2,

profiling does not decrease the amount of sample material available for DNA profiling because during coextraction, the RNA is isolated from a portion of the DNA extract that is generally discarded when DNA is the only target of the extraction process.

Several RNA and RNA/DNA coextraction kits are commercially available and suitable for the analysis of forensic stains [7, 12–14]. In the following, we describe the use of two RNA and one RNA/DNA coextraction kits, one of which is designed for the isolation of total RNA from minimal amounts of biological material. DNase treatment (on-column or post-RNA purification) is recommended to avoid primer and nucleotide titration by contaminating DNA.

Several reverse transcription PCR (RT-PCR) kits are available and their suitability for use with forensic biological samples has been demonstrated [7, 12–14]. One method is described in detail in the Materials and Methods section. An RT minus control (no reverse transcriptase added) can be used to detect possible DNA contamination.

Quantitation of the RNA derived from sample types likely to be found in forensic casework has proven to be problematic. To date, there are no published methods for human specific RNA quantitation that give an accurate prediction of the success of a subsequent mRNA profile from a forensic type sample; therefore, no RNA quantification methodology will be discussed in this chapter.

Designing mRNA profiling PCR multiplexes is more challenging than assays where genomic DNA is the target due to the markers for a single body fluid/tissue having different susceptibilities to degradation and copy numbers within a cell, mRNA markers showing variable expression between samples of the same type and mRNA expression not being absolutely restricted to the target body fluid/tissue. In an assay where multiple markers are included for a single target body fluid/tissue, the peak heights for these markers in an mRNA profile generated from a single source body fluid are generally not balanced and in fact can vary widely (Fig. 1). It is advisable to use multiple markers per target body fluid/tissue to decrease the possibility of a false positive result due to unexpected amplification of a marker in a nontarget body fluid/tissue. A list of suitable RNA markers for the identification of body fluids is given in Table 1.

Guidelines for interpretation of mRNA profiles must be developed for each assay in order to avoid false positives due to occasional nontarget amplification of markers. A large number of control samples of each target body fluid should be used to determine the behavior of each marker in both the target body fluid/tissue and in nontarget samples under the conditions used for each

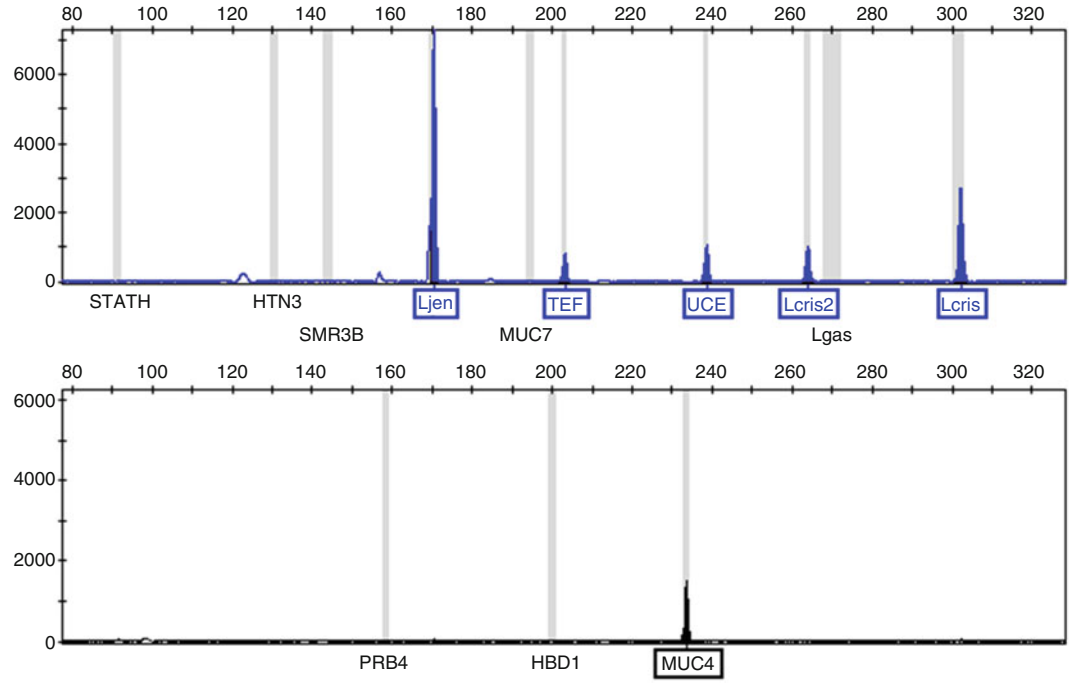


Fig. 1 An RNA profile generated from a vaginal swab. This multiplex consists only of saliva (STATH, HTN3, SMR3B, Muc7, PRB4), CVF (Ljen, Lcris2, Lcris, Lgas, HBD1, Muc4), and housekeeping (UCE, TEF) markers [1]. The saliva score for this sample, as discussed in the introduction, is 0 because none of the saliva markers have amplified. The CVF score for this sample would be the sum of the predetermined values for Ljen, Lcris2, Lcris, and MUC4. The scores for HBD1 and Lgas are 0. If the CVF score is $\geq$ the predetermined minimum score for CVF, the sample is determined to be positive for this body fluid

assay. These results can then be used to devise an appropriate interpretation system. There are a number of published methods for interpretation of these profiles [1, 15, 16]. Our approach for assays utilizing multiple markers per body fluid is to use the results from a panel of control samples of each target body fluid to generate positive predictive values [PPV = true positives/(true positives + false positives)] for each marker [1]. Using this system, markers that are rarely observed in nontarget samples receive scores close to 1 and markers that show less biased expression in their respective target body fluids receive lower scores. If a marker is observed in a profile, that marker receives its preallotted score. If that marker is not amplified, it is given a score of 0. The values for markers targeting the same body fluid are added. This is done for each body fluid in the assay. Using the results from the control samples, minimum scores that must be achieved in order for a positive result are set for each body fluid/tissue.

Table 1
Primer sequences for body fluid RNA and housekeeping gene markers

Target body fluid	Marker [gene name]	Primer sequence	Size (bp)	Reference
Semen	KLK3 [kallikrein-related peptidase 3]	F: 5′ TTGTCTTCCTCACCCTGTCC	100	[1]
		R: 5′ CAGGGTTGGGAATGCTTCT		[1]
	PRM1 [protamine 1]	F: 5′ GCCAGGTACAGATGCTGTCGCAG	153	[2]
		R: 5′ TTAGTGTCTTCTACATCTCGGTCT		[2]
	SEMG1 [semenogelin I]	F: 5′ ACAACGACCGAAACCCATTA	158	[1]
		R: 5′ GTGTCATCCATGGACCAAGA		[3]
	TGM4 [transglutaminase 4]	F: 5′ TGAGAAAGGCCAGGGCG	215	[1]
		R: 5′ AATCGAAGCCTGTCACACTGC		[1]
	PRM2 [protamine 2]	F: 5′ GTGAGGAGCCTGAGCGAACGC	294	[2]
		R: 5′ TTAGTGCCTTCTGCATGTTCTCTTC		[2]
Saliva	STATH [statherin]	F: 5′ TTTGCCTTCATCTTGGCTCT	93	[4]
		R: 5′ CCCATAACCGAATCTTCCAA		[4]
	HTN3 [histatin 3]	F: 5′ GCAAAGAGACATCATGGGTA	134	[5]
		R: 5′ GCCAGTCAAACCTCCATAATC		[5]
	SMR3B [submaxillary gland androgen regulated protein 3B]	F: 5′ GCAACTGAAAGGATGAAATCACTGACTTG	147	[1]
		R: 5′ AAATCCTGGGCCAAAAGGTTGAGGAG		[1]
	PRB4 [proline-rich protein BstN1 subfamily 4]	F: 5′ GCAGGCCACCCAGACCTGCCCAGG	160	[1]

		R: 5′ GGTATATTAAAGTTAGAGCTATGATGACCTTG		[1]
	MUC7 [mucin 7]	F: 5′ GTTTGTGTGCATCTGTGCACTGAGTG	197	[1]
		R: 5′ GCCTACAGCGTTTGTGCAGACATTTATAGG		[1]
CVF	Ljen [16 s ribosomal RNA]	F: 5′ AAGTCGAGCGAGCTTGCCTATTGAAAT	171	[1]
		R: 5′ CGCCTTTTAAACTTCTTTCATGCGAAAGTAGC		[1]
	HBD1 [beta defensin 1]	F: 5′ CCTGGGTGTTGCCTGCCAGTCGC	200	[1]
		R: 5′ CAGGTGCCTTGAATTTT**GGT**[a]		[6]
	MUC4 [mucin 4]	F: 5′ GGACCACATTTTATCAGGAA	235	[5]
		R: 5′ TAGAGAAACAGGGCATAGGA		[5]
	Lgas [16S–23S intergenic spacer region]	F: 5′ ATGATGGAGAGTGCGAGAGC	311	[7]
		R: 5′ CCGGATCATTGCTTACTTAC		[7]
	Lcris [16S–23S intergenic spacer region]	F: 5′ GAGAGCAGGAATGCTAAGAG	292	[7]
		R: 5′ CCGGATCATTGCTTACTTAC		[7]
	MYOZ1	F: 5′ GGGTTGGTGAGACAGGATCA	81	[8]
		R: 5′ TCCCATGGGGAAATATAGGT		[8]
	CYP2B7P1	F: 5′ TCCTTTCTGAGGTTCCGAGA	198	[8]
		R: 5′ TTTCCATTGGCAAAGAGCAT		[8]
Blood	HBB [hemoglobin subunit beta]	F: 5′ GCACGTGGATCCTGAGAAC	61	[4]
		R: 5′ ATGGGCCAGCACACAGAC		[4]
	ALAS2 [delta-aminolevulinate synthase]	F: 5′ GCCAATGACTCTACTCTCTTCACCTTGGC	81	[1]

(continued)

Table 1
(continued)

Target body fluid	Marker [gene name]	Primer sequence	Size (bp)	Reference
		R: 5′ GGAAGCATGGTTGCCTGCGTCTGAG		[1]
	PRF1 [perforin]	F: 5′ GGACCAGTACAGCTTCAGCACTGACACG	102	[1]
		R: 5′ CCGAGGGCCCTCTTGAAGTCAGGGTG		[1]
	GlycoA [glycophorin A]	F: 5′ GGGTGATGGCTGGTGTTATTGGAACGATC	115	[1]
		R: 5′ GGCACGTCTGTGTCAGGTGAGGGGAG		[1]
	PF4 [platelet factor 4]	F: 5′ GGTCCGTCCCAGGCACATCACCAGC	124	[1]
		R: 5′ CAGCGGGGCTTGCAGGTCCAAG		[1]
	SPTB [erythrocytic spectrin beta chain]	F: 5′ AGGATGGCTTGGCCTTTAAT	150	[5]
		R: 5′ GGTCGAGGAGCGGGATGATGCC		[1]
	PBGD [hydroxymethylbilane synthase]	F: 5′ TGGATCCCTGAGGAGGGCAGAAG	177	[6]
		R: 5′ TCTTGTCCCCTGTGGTGGACATAGCAAT		[6]
MB	MMP11 [metallopeptidase 11]	F: 5′ GGTGCCCTCTGAGATCGAC	92	[4]
		R: 5′ TCACAGGGTCAAACTTCCAGT		[4]
	MSX1 [msh homeobox 1]	F: 5′ CCCCGTGGATGCAGAGCCCCCG	96	[1]
		R: 5′ GCTTACGGTTCGTCTTGTGTTTGCGGAG		[1]
	SFRP4 [secreted frizzled-related protein 4]	F: 5′ GCGACGAGCTGCCTGTCTATGACC	136	[1]
		R: 5′ CAGTCAACATCAAGAGGCCTTTCCTGTAC		[1]
	LEFTY2 [left-right determination factor 2]	F: 5′ GCCCACGTGAGGGCCCAGTATGTAGT	130	[1]

		R: 5′ GGTGTGTGCTGGCCTCCGACGC		[1]
	MMP10 [metallopeptidase 10]	F: 5′ TGCCATCAGCACTCTGAGGGGAGAATAT	143	[1]
		R: 5′ CATATGCAGCATCCAAATATGATGGAAGAGAG		[1]
	MMP7 [metallopeptidase 7]	F: 5′ CATGAGTGAGCTACAGTGGGAACAGGC	161	[1]
		R: 5′ CTATGACGCGGGAGTTTAACATTCCAG		[1]
	Hs202072[b] [uncharacterized LOC100505776]	F: 5′ CTCTGACGTTTCATCTCCTGTTTATGAGGAC	168	[1]
		R: 5′ GATGAAAATGTCACCACTCTAATTCAGGCC		[1]
Housekeeping	TEF[c] [transcription elongation factor SII]	F: 5′ TGGGCCATCAACTGAGAAAGA	206	[9]
		R: 5′ TCTCCCTACACTTCAACTGCACA		[9]
	UCE [E2 ubiquitin conjugating enzyme UbcH5B]	F: 5′ AATGATCTGGCACGGGACC	241	[9]
		R: 5′ ATCGTAGAATATCAAGACAAATGCTGC		[9]
	B2M [Beta-2 microglobulin]	F: 5′ GGCATTCCTGAAGCTGACA	120	[7]
		R: 5′ AAACCTGAATCTTTGGAGTACG		[7]
	UBC [Ubiquitin C)	F: 5′ GGGTCGCAGTTCTTGTTTGT		[7]
		R: 5′ TCCAGCAAAGATCAGCCTCT		[7]

[a]Bases in bold are different from the cited primers
[b]UGID: 162601 (http://www.ncbi.nlm.nih.gov/unigene)
[c]In addition to TEF cDNA, this primer set also amplifies a TEF pseudogene that is identical in size to the cDNA

2 Equipment and Materials

All protocols listed for use with kits are based on the instructions supplied by the manufacturers.

2.1 General Supplies and Equipment

1. Vortex mixer.
2. Ultrasonic bath (e.g., Ultrawave U300H).
3. Fume hood.
4. Thermomixer/heat block.
5. Microcentrifuge.
6. Thermocycler.
7. 3130 or 3500 Genetic Analyzer (Life Technologies).
8. RNaseZap® wipes and spray (Ambion®) (*see* **Note 1**).
9. Microcentrifuge tubes (e.g., DNA LoBind® (Eppendorf)) (*see* **Note 2**).
10. Spin baskets.
11. Forceps.

2.2 RNA Extraction with the RNeasy® Plus Mini Kit (QIAGEN GmbH) and RNA/DNA Coextraction with the AllPrep® DNA/RNA Mini Kit (QIAGEN GmbH)

1. RNeasy® Plus Mini kit (QIAGEN GmbH) or AllPrep® DNA/RNA Mini kit (QIAGEN GmbH). The required buffers are supplied with both kits. 96–100 % ethanol must be added as per the manufacturer's instructions to the AW1, AW2, and RPE buffers prior to use.
2. 14.3 M β-mercaptoethanol (βME) is required for addition to RLT Plus buffer. A working solution of this buffer contains 10 μl βME per 1 ml of RLT Plus buffer. Add βME to an aliquot of the RLT Plus buffer rather than the entire bottle as the RLT Plus buffer/βME mixture is only stable at room temperature for 1 month.
3. 70 % ethanol (dilute ethanol in RNase-free H_2O).
4. Either on-column or post-RNA extraction DNAase treatment is recommended (*see* **Note 3**). The RNase-free DNase set (QIAGEN GmbH) can be used for on-column DNA digestion. Make DNase solution according to the manufacturer's instructions. Alternatively, the TURBO DNA-free™ Kit (Life Technologies) can be used following RNA extraction (*see* Subheading 2.4).

2.3 RNA Extraction with Arcturus® PicoPure® RNA Isolation Kit (Life Technologies)

1. All necessary reagents are supplied with the kit.

2.4 DNase Treatment with the TURBO DNA-Free Kit™ (Life Technologies)

1. All necessary reagents are supplied with the kit.

2.5 Reverse Transcription

1. SuperScript® III Reverse Transcriptase (Life Technologies). The reverse transcriptase is supplied with 5× first-strand buffer and 0.1 M DTT.
2. RNaseOUT™ (40 U/μl) (Life Technologies).
3. Random primers (50 ng/μl) (Life Technologies).
4. RNase-free H_2O.
5. 10 mM dNTP mix (QIAGEN GmbH).
6. Freezer or ice.

2.6 PCR of cDNA

1. QIAGEN Multiplex PCR kit (QIAGEN GmbH). The kit is supplied with 2× QIAGEN Multiplex PCR Master Mix, 5× Q-solution, and RNase-free H_2O.
2. Primer pairs for RNA markers amplification (Table 1). One primer in each pair should be 5′ fluorescently labeled for detection during capillary electrophoresis (*see* **Note 4**).

2.7 Post-PCR Purification

1. MinElute® PCR purification kit (QIAGEN GmbH). This kit contains MinElute spin columns and buffers PB, PE, and EB. Ethanol (96–100 %) must be added to PE buffer before use.

2.8 mRNA Profiling: Capillary Electrophoresis

1. GeneScan® 500 Liz™ size standard or GeneScan® 400HD ROX™ size standard (Life Technologies) (*see* **Note 4**).
2. Hi-Di™ formamide (Life Technologies).
3. POP-4™ polymer (Life Technologies).
4. GeneMapper® software (Life Technologies).

3 Methods

Great care must be taken to avoid introducing RNases to the samples. Wipe bench area, equipment (including pipettes and forceps), and reagent bottles with RNaseZap® prior to use (*see* **Note 1**). Change gloves frequently and wipe/spray gloves with RNaseZap® during RNA extraction and reverse transcription setup. Use RNase-free pipette tips and microcentrifuge tubes.

Due to its toxicity, a fume hood should be used when working with βME.

3.1 RNA/DNA Coextraction with the AllPrep® RNA/DNA Mini Kit (QIAGEN GmbH)

This kit can be used to purify RNA and DNA from a single sample. If DNA extraction is not required, the RNeasy® Plus Mini kit (QIAGEN GmbH; Subheading 3.2) or, in the case of very low template samples, the Arcturus® PicoPure® RNA Isolation Kit (Life Technologies; Subheading 3.3) may be more suitable.

1. Before starting the extraction preheat the required amount of EB buffer (50 µl/sample) to 70 °C.
2. For each sample, add 400 µl RLT Plus/βME mixture to a 1.5 ml microcentrifuge tube.
3. Transfer the sample (swab head or sample cutting) in the microcentrifuge tube.
4. Vortex the sample.
5. Incubate for 10 min at room temperature. During the incubation, vortex the sample for approximately 2 min, then sonicate for 5 min followed by vortexing for 1–2 min. If the stain has not solubilized after 10 min, additional sonication and vortexing steps can be performed.
   Alternatively, vortex the sample for 1–2 min and incubate for 3 h at 56 °C.
6. Briefly centrifuge the sample.
7. Transfer swab head/sample cutting to a spin basket, place in microcentrifuge tube and then centrifuge at maximum speed for 5 min.
8. Remove the spin basket/sample (*see* **Note 5**). Transfer the eluate to an AllPrep® DNA spin column. Care should be taken during pipetting to avoid air bubbles as the solution will foam. If the full volume of eluate cannot be added to the spin column due to foaming, gently tapping the side of the tube will cause the air bubbles to dissipate.
9. Centrifuge for 30 s at 11,000 × *g*. No fluid should remain on the spin column membrane. If there is fluid, recentrifuge the sample.
10. Keep the collection tube containing the eluate for the RNA extraction (**steps 12–34**).
11. Place the AllPrep® DNA spin column in a new 2 ml collection tube. This spin column will be used for DNA purification (**steps 35–45**).

    RNA extraction

12. To the eluate generated in **step 9** above, add 400 µl of 70 % ethanol. Mix with a pipette.
13. Avoiding air bubbles, transfer the eluate/ethanol to a RNeasy® MinElute spin column.
14. Carefully close the tube and centrifuge for 15 s at 11,000 × *g*.

15. Discard the flow-through and place the spin column back in the collection tube.
16. For on-column DNase digestion follow **steps 18–22**. If the samples are not to undergo on-column DNase digestion, add 700 µl of buffer RW1 and go to **step 23**.
17. Add 350 µl RW1 buffer to the spin column.
18. Centrifuge for 15 s at 11,000 × *g*.
19. Add 80 µl of DNase solution (10 µl DNase I stock solution (stored at 4 °C) and 70 µl RDD buffer) directly to the membrane. Mix gently by tapping the tube. Do not vortex or mix with a pipette (*see* **Note 6**).
20. Incubate for 15 min at RT (20–30 °C).
21. Add 350 µl of RW1 buffer to the spin column.
22. Centrifuge for 15 s at 11,000 × *g*.
23. Discard the flow-through and place spin column back in the collection tube.
24. Add 500 µl of RPE buffer to the spin column. Make sure that ethanol has been added to the RPE buffer.
25. Centrifuge for 15 s at 11,000 × *g*.
26. Discard the flow-through and place spin column back in the collection tube.
27. Add 500 µl of RPE buffer to the spin column and centrifuge for 2 min at 11,000 × *g*.
28. Place spin column in a new 2 ml collection tube. Discard the flow-through and collection tube.
29. Centrifuge the spin column for 1 min at maximum speed. It is important that the ethanol does not carry over into the RT-PCR reaction (*see* **Note 7**).
30. Place the spin column in a 1.5 ml collection tube.
31. Without touching the spin column membrane with the pipette tip, pipette 30 µl of RNase-free H_2O directly to the membrane.
32. Incubate 1–5 min at room temperature and centrifuge for 1 min at 11,000 × *g* to elute the RNA.
33. Optional: for increased yield, place the eluate onto the spin column membrane and repeat **step 33**.
34. The eluate contains the purified RNA. The RNA can now be used for cDNA synthesis (Subheading 3.5) or DNase treated if this has not been done during RNA purification (Subheading 3.4).

DNA extraction

35. Add 500 µl of AW1 buffer to the DNA spin column from **step 12**. Make sure that ethanol has been added to the buffer.

36. Centrifuge for 15 s at 11,000×*g* and discard the flow-through.
37. Add 500 µl of AW2 buffer to the spin column. Make sure that ethanol has been added to the buffer.
38. Centrifuge for 2 min at maximum speed.
39. Discard the flow-through and centrifuge the spin column for 1 min at 13,000 rpm.
40. Place the DNA spin column into a 1.5 ml collection tube.
41. To elute the DNA from the spin column, add 50 µl EB buffer (preheated to 70 °C) directly to the membrane.
42. Incubate for 1 min at room temperature.
43. Centrifuge for 1 min at 11,000×*g*.
44. Optional: for increased DNA yield, place the eluate onto the spin column membrane and repeat **steps 43–44**.
45. Discard the spin column. The eluate contains the purified DNA.

3.2 RNA Extraction with the RNeasy® Plus Mini Kit (QIAGEN GmbH)

This kit can be used when coextraction of DNA is not required. For samples thought to contain very low quantities of biological material, the Arcturus® PicoPure® RNA Isolation Kit (Life Technologies; Subheading 3.3) may be more suitable.

1. For each sample, add 400 µl RLT Plus/βME mixture to a 1.5 ml microcentrifuge tube.
2. Transfer the sample (swab head or sample cutting) into the microcentrifuge tube.
3. Vortex the sample.
4. Incubate for 10 min at room temperature.
5. During the incubation, vortex the sample for approximately 2 min, then sonicate for 5 min followed by vortexing for 1–2 min. If stain has not solubilized after 10 min, additional sonication and vortexing steps can be performed.
   Alternatively, vortex the sample for 1–2 min and incubate for 3 h at 56 °C.
6. Briefly centrifuge the sample.
7. Transfer swab head/cutting to a spin basket.
8. Centrifuge at maximum speed for 5 min.
9. Remove the spin basket (*see* **Note 5**). Transfer the eluate to a genomic DNA (gDNA) eliminator spin column. Care should be taken during pipetting to avoid air bubbles as the solution will foam. If the full volume of eluate cannot be added to the spin column due to foaming, gently tapping the side of the tube will cause the air bubbles to dissipate.

10. Centrifuge for 30 s at 11,000 × *g*. No fluid should remain on the spin column membrane. If there is fluid, recentrifuge the sample.
11. Discard the spin column and add 400 µl of 70 % ethanol to the eluate. Mix with a pipette.
12. Avoiding air bubbles, transfer the eluate/ethanol to a RNeasy® spin column.
13. Carefully close the tube and centrifuge for 15 s at 11,000 × *g*.
14. Discard the flow-through and place the spin column back in the collection tube.
15. For on-column DNase digestion, follow **steps 16–19**. If the samples are not to undergo on-column DNase digestion, add 700 µl of buffer RW1 and go to **step 20**.
16. Add 350 µl RW1 buffer to the spin column and centrifuge for 15 s at 11,000 × *g*.
17. Add 80 µl of DNase solution (10 µl DNase I stock solution (stored at 4 °C) and 70 µl RDD buffer) directly to the membrane. Mix gently by tapping the tube. Do not vortex or mix with a pipette (*see* **Note 6**).
18. Incubate for 15 min at RT (20–30 °C).
19. Add 350 µl of RW1 buffer to the spin column.
20. Centrifuge for 15 s at 11,000 × *g*.
21. Discard the flow-through and place the spin column back in the collection tube.
22. Add 500 µl of RPE buffer to the spin column. Make sure that ethanol has been added to the RPE buffer.
23. Centrifuge for 15 s at 11,000 × *g*, discard the flow-through, and place the spin column back in the collection tube.
24. Add 500 µl of RPE buffer to the spin column.
25. Centrifuge for 2 min at 11,000 × *g*.
26. Place spin column in a new 2 ml collection tube. Discard the flow-through and used collection tube.
27. Centrifuge the spin column for 1 min at maximum speed. It is important that the ethanol does not carry over into the RT-PCR reaction (*see* **Note 7**).
28. Place column in a new 1.5 ml collection tube.
29. Without touching the spin column membrane with the pipette tip, pipette 30 µl of RNase-free H_2O directly to the membrane.
30. Incubate 1–5 min at room temperature and centrifuge for 1 min at 11,000 × *g* to elute the RNA.

31. Optional: for increased RNA yield, place the eluate onto the membrane and repeat **step 30**.
32. The eluate contains the purified RNA. The RNA can now be used for cDNA synthesis (Subheading 3.5) or DNase treated if this has not been done during RNA purification (Subheading 3.4).

3.3 RNA Extraction with the Arcturus® PicoPure® RNA Isolation Kit (Life Technologies)

This kit can be used to isolate total RNA from samples thought to contain minimal amounts of biological material. The kit is designed to isolate total RNA from samples containing a single cell.

1. Precondition the purification column by pipetting 250 μl of Conditioning Buffer (CB) onto the purification column filter membrane and incubate for 5 min at room temperature.
2. Centrifuge at 13,000 rpm for 1 min.
3. Place the sample (swab head or sample cutting) in a 1.5 ml microcentrifuge tube.
4. Add 100 μl Extraction Buffer (XB) directly to the swab head/sample cutting.
5. Incubate at 42 °C for 30 min.
6. Using clean/RNase-free forceps, take the XB soaked swab/sample cutting out of the extraction solution and place it into a spin basket. Place the basket in the extraction tube and centrifuge at 13,000 rpm for 5 min. Discard the spin basket and the dry swab (*see* **Note 5**).
7. Pipette 100 μl of 70 % ethanol into the cell extract and mix well by pipetting up and down.
8. Pipette the cell extract and ethanol mixture into the preconditioned purification column.
9. Centrifuge for 2 min at 1000 rpm, immediately followed by centrifugation at 13,000 rpm for 30 s.
10. Pipette 100 μl of wash buffer 1 (W1) into the purification column and centrifuge for 1 min at 9000 rpm.
11. Pipette 100 μl wash buffer 2 (W2) into the purification column and centrifuge for 1 min at 9000 rpm.
12. Pipette 100 μl wash buffer 2 (W2) into the purification column and centrifuge for 2 min at 13,000 rpm. If wash buffer remains, recentrifuge for 1 min at 13,000 rpm.
13. Transfer the purification column to a new 0.5 ml microcentrifuge tube.
14. Pipette 12 μl of Elution buffer (EB) or DNase/RNase-free H_2O directly onto the membrane of the purification column.
15. Centrifuge for 1 min at 3500 × *g* followed by a 1 min centrifugation at 13,000 rpm.

16. Discard the purification column. The RNA is ready for DNase treatment (Subheading 3.4) or reverse transcription (Subheading 3.5).

3.4 Post-RNA Purification DNase Treatment with the TURBO DNA-Free Kit™ (Life Technologies)

This DNase treatment method allows for more thorough removal of contaminating DNA than on-column DNase treatment (*see* **Note 3**). This post-purification treatment is particularly important in samples that are thought to contain large amounts of biological material.

1. Place 30 µl of RNA into a 0.5 ml microcentrifuge tube.
2. Add 3 µl 10× TURBO DNase buffer and 1 µl TURBO DNase to the sample.
3. Incubate 20–30 min at 37 °C.
4. Add 3.4 µl resuspended Inactivation Reagent and mix well.
5. Incubate 5 min at room temperature, mixing occasionally to redisperse the Inactivation Reagent.
6. Centrifuge for 1.5 min at 10,000 × *g*.
7. Without transferring the pelleted DNase Inactivation Reagent, pipette the supernatant (about 30–32 µl) into a new 0.5 ml tube (*see* **Note 8**). It may not be possible to recover all of the RNA.

3.5 cDNA Synthesis

During reverse transcription, RNA is used as a template for the synthesis of complementary DNA (cDNA) strands. The resulting cDNA can then be used as template in PCR.

For each RNA sample, set up a reverse transcriptase positive (RT+) and RT– (no enzyme) reaction. The RT– reaction is a control for the presence of gDNA in the sample.

The enzymes RNaseOUT™ and Superscript™ III reverse transcriptase should be kept on ice or in the freezer until immediately before use and then returned to the freezer immediately after use.

1. Set a heat block to 65 °C and a thermomixer to 50 °C. If the mixer is programmable to multiple temperatures, set it to 50 °C for 45 min followed by 70 °C for 15 min (alternatively, a thermocycler can be used).
2. Thaw the random primers (50 ng/µl), 10 mM dNTPs, 0.1 M DTT, and 5× first-strand buffer.
3. (a) If the sample has not been DNase treated or has undergone on-column DNase treatment during RNA purification, transfer 25 µl of RNA into a 1.5 ml microcentrifuge tube.

   (b) If the RNA has been treated with the TURBO DNA-free Kit™, transfer up to 16.8 µl of RNA into a 1.5 ml microcentrifuge tube and add RNase-free H_2O to make a final volume of 25 µl (*see* **Note 9**).

4. Add 2 µl of random primers and 2 µl of dNTPs to the sample.
5. Incubate for 5 min at 65 °C to denature the RNA.
6. During the incubation, prepare the reverse transcription mix.
   8.4 µl 5 × first-strand buffer.
   2 µl 0.1 M DTT.
   2 µl RNaseOUT™ (40 U/µl).
7. After incubation, place the sample either on ice or in the freezer for 1–2 min. Do not allow the sample to freeze.
8. Briefly centrifuge the sample.
9. Add the reverse transcription mix to the sample and mix with a pipette.
10. To a new 0.5 ml microcentrifuge tube or 0.2 ml PCR tube (depending on the thermomixer/thermocycler to be used), add 1 µl of Superscript™ III reverse transcriptase (200 U/µl). This will be the RT+ reaction.
11. To a second 0.5 ml microcentrifuge tube or 0.2 ml PCR tube (depending on the thermomixer/thermocycler to be used), add 1 µl of RNase-free H_2O. This will be the RT– reaction.
12. Pipette 20 µl of reverse transcription mix per sample into the corresponding RT+ and RT– tubes.
13. Gently mix the sample and then briefly centrifuge.
14. Incubate at room temperature for 5 min.
15. Incubate at 50 °C for 45 min followed by enzyme inactivation at 70 °C for 15 min.
16. Briefly centrifuge the sample.
17. Leave the sample at room temperature if the PCR is going to be set up immediately; otherwise store samples at –20 °C.

3.6 cDNA PCR

This process will amplify specific, body fluid markers that are present in the sample.

Multiplexes should be designed so that the PCR amplicons do not overlap in size regardless of whether they are labeled with different fluorescent dyes (*see* **Note 10**).

Due to the possibility of marker amplification in nontarget body fluids and variability of marker expression in target body fluids, multiple markers should target each body fluid in the assay(s). Housekeeping markers can be included in each assay as positive controls (*see* **Note 11**).

When multiple markers are used for each body fluid, it may be necessary to design more than one multiplex. In this circumstance, all samples should be run with all multiplexes so that the results can be interpreted in context.

Preliminary work should be done to determine the optimal primer concentration for each marker in the assay. Generally, a

concentration of 0.2 μM per primer in the PCR is a good starting point for making this determination.

PCR should be done on both the RT+ and RT– reaction for each sample.

1. If frozen, thaw cDNA (RT+) and RT– control, mix and centrifuge.
2. Thaw the QIAGEN multiplex mix, primer multiplex mix(es), and H_2O. Vortex the thawed multiplex and primer mixes before use.
3. The amount/volume of cDNA added can be varied and should be optimized for each PCR. Add equal volumes of cDNA and RT– to their respective PCRs.
4. The following is an example of a 10 μl PCR:
   - 5 μ 2× QIAGEN multiplex mix.
   - 3 μl cDNA.
   - 1 μl primer mastermix.
   - 1 μl H_2O.
5. Amplification conditions should be optimized for each primer multiplex and thermocycler. The following is an example:

95 °C for 15 min	
30–35 cycles of:	94 °C for 15 s
	58 °C for 90 s
	72 °C for 1 min
72 °C for 20 min	
15 °C hold	

6. Once thermocycling is completed, the PCR products can be stored at 4 °C (short term) or at –20 °C.

3.7 Post-PCR Purification with the MinElute® PCR Purification Kit (QIAGEN)

1. Add 125 μl buffer PB to 25 μl PCR reaction mix.
2. Transfer the sample to a MinElute® column and centrifuge for 1 min at 13,000 rpm. Discard the flow-through.
3. Add 750 μl Buffer PE to the MinElute® column and centrifuge for 1 min at 13,000 rpm. Discard the flow-through.
4. Centrifuge the column for an additional 1 min at 13,000 rpm. Place the MinElute® column in a clean 1.5 ml microcentrifuge tube.
5. Add 10–30 μl Buffer EB or H_2O to the center of the membrane. Let the column stand for 1 min at room temperature and then centrifuge for 1 min at 13,000 rpm to elute the DNA. Discard the column.

3.8 mRNA Profiling: Capillary Electrophoresis

1. For capillary electrophoresis of the PCR product, add 1 μl of PCR product to 8.5 μl of Hi-Di™ formamide (Life Technologies) and 0.5 μl of either GeneScan® 500 Liz™ size standard or GeneScan® 400HD ROX™ size standard (Life Technologies) (*see* **Note 4**).
2. For electrophoresis on a 3130xl Genetic Analyzer (Life Technologies) the settings are:
   - 10 s injection time.
   - 3 kV injection voltage.
   - 15 kV run voltage.
   - 60 °C.
   - 23.5 min run time (adjust time depending on the length of the amplicons in assay).
3. Following electrophoresis the raw data are analyzed using GeneMapper® software (Life Technologies). Marker scoring thresholds should be determined for each multiplex.

An example of an mRNA profile generated from a vaginal swab is shown in Fig. 1. The described scoring system (Subheading 1) was applied to this stain for profile interpretation.

4 Notes

1. RNaseZap® solution can corrode metals. Avoid extended periods of contact with metal surfaces and rinse all items thoroughly with H_2O after use to remove RNaseZap®.
2. DNA LoBind® tubes are designed to decrease the absorption of nucleic acids to the tube during sample processing and storage.
3. Based on our experience, the RNase-free DNase Set for on-column DNase digestion does not completely remove contaminating DNA in samples that contain relatively large amounts of cellular material. The TURBO DNA-free Kit™ efficiently digests any DNA carried through the RNA extraction process.
4. The fluorescent dyes are 5-FAM™, JOE™, NED™, and ROX™ (Dye Set F) or 6-FAM™, VIC®, NED™, PET®, and LIZ® (Dye Set G5). The size standard, GeneScan® 500 Liz™ or GeneScan® 400HD ROX™, is chosen based on which dyes are used to label the PCR primers.
5. The spin basket and swab head/sample cutting can either be discarded or, if required, stored frozen.
6. The DNase I enzyme is susceptible to physical denaturation.
7. The presence of ethanol can inhibit the reverse transcription reaction.

8. The presence of DNase Inactivation Reagent affects subsequent enzymatic reactions due to its ability to sequester divalent cations.
9. The manufacturer recommends that the TURBO DNA-free™ treated RNA comprises ≤40 % of the final reverse transcription volume because carry-over of reagents from the DNase digestion could interfere with the reaction.
10. The level of amplification of each marker can vary widely among samples comprised of the same body fluid. For this reason, overamplified peaks and the resulting pull-up in other dye channels are not uncommon in these profiles.
11. To date, no housekeeping marker has been described that amplifies equally well in semen, blood, MB, saliva, and CVF samples. Detection of these markers is particularly poor in saliva and semen. In these sample types, the housekeeping markers may not be amplified while all of their respective body fluid markers are detected.

References

1. Roeder AD, Haas C (2013) mRNA profiling using a minimum of five mRNA markers per body fluid and a novel scoring method for body fluid identification. Int J Legal Med 127:707–721
2. Steger K, Pauls K, Klonisch T, Franke FE, Bergmann M (2000) Expression of protamine-1 and –2 mRNA during human spermiogenesis. Mol Hum Reprod 6:219–225
3. Haas C, Muheim C, Kratzer A, Bär W, Maake C (2009) mRNA profiling for the identification of sperm and seminal plasma. Forensic Sci Int Genet Suppl Ser 2:534–535
4. Haas C, Klesser B, Maake C, Bär W, Kratzer A (2009) mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. Forensic Sci Int Genet 3:80–88
5. Alvarez M, Juusola J, Ballantyne J (2004) An mRNA and DNA co-isolation method for forensic casework samples. Anal Biochem 335:289–298
6. Juusola J, Ballantyne J (2005) Multiplex mRNA profiling for the identification of body fluids. Forensic Sci Int 152:1–12
7. Haas C, Hanson E, Anjos MJ, Ballantyne K, Banemann R et al (2014) RNA/DNA co-analysis from human vaginal and menstrual blood stains – results of a fourth and fifth collaborative EDNAP exercise. Forensic Sci Int Genet 8(1):203–212
8. Hanson EK, Ballantyne J (2013) Highly specific mRNA biomarkers for the identification of vaginal secretions in sexual assault investigations. Sci Justice 53:14–22
9. Fleming RI, Harbison SA (2010) The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids. Forensic Sci Int Genet 4:244–256
10. Hanson E, Haas C, Jucker R, Ballantyne J (2011) Specific and sensitive mRNA biomarkers for the identification of skin in 'touch DNA' evidence. Forensic Sci Int Genet 6:548–558
11. Visser M, Zubakov D, Ballantyne KN, Kayser M (2011) mRNA-based skin identification for forensic applications. Int J Legal Med 125:253–263
12. Haas C, Hanson E, Anjos MJ, Banemann R, Berti A et al (2013) RNA/DNA co-analysis from human saliva and semen stains – results of a third collaborative EDNAP exercise. Forensic Sci Int Genet 7:230–239
13. Haas C, Hanson E, Anjos MJ, Bär W, Banemann R (2012) RNA/DNA co-analysis from blood stains – results of a second collaborative EDNAP exercise. Forensic Sci Int Genet 6:70–80
14. Haas C, Hanson E, Bär W, Banemann R, Bento AM et al (2011) mRNA profiling for the identification of blood – results of a collaborative EDNAP exercise. Forensic Sci Int Genet 5:21–26
15. Lindenbergh A, de Pagter M, Ramdayal G, Visser M, Zubakov D et al (2012) A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces. Forensic Sci Int Genet 6:565–577
16. Lindenbergh A, Maaskant P, Sijen T (2013) Implementation of RNA profiling in forensic casework. Forensic Sci Int Genet 7:159–166

Chapter 3

mRNA Profiling for Vaginal Fluid and Menstrual Blood Identification

Joanna Jakubowska, Agnieszka Maciejewska, and Ryszard Pawłowski

Abstract

Recently mRNA profiling has been widely proposed as a universal tool for biological fluids identification. Here, we describe a test for vaginal fluid identification that combines detection of five markers: vaginal mRNAs and Lactobacilli in end-point PCR reaction. The test detects the following transcripts: *HBD1* (Human beta-defensin 1), *MUC4* (Mucin 4), *MMP11* (Matrix metalloproteinase 11), housekeeping gene *G6PDH* (glucose 6-phosphate dehydrogenase), and the 16S–23S rRNA intergenic spacer regions of *L. crispatus* and *L. gasseri/L. johnsonii*. Simultaneous analysis of five vaginal markers and a housekeeping gene ensures high specificity and reliability in the detection of vaginal material, which could not be obtained using detection of a single marker.

Key words Vaginal fluid, Menstrual blood, mRNA profiling, Lactobacillus sp., Multiplex PCR, Forensic genetics

1 Introduction

Vaginal fluid and menstrual blood identification are highly important in forensic biology, especially in the case of sexual crimes. Unfortunately, there are no confirmatory tests for detection of vaginal secretions among classical methods [1]. Recently mRNA profiling has been widely proposed as a universal tool for biological fluid detection. Every tissue and body fluid posses its own and specific mRNA profile, so it is possible to identify biological material using specific mRNA markers. Application of the PCR technique increases the sensitivity of the test in comparison to classical confirmatory methods. RNA analysis consists of the following steps: RNA isolation, quantification, reverse trancription, amplification of particular transcripts, and detection of amplicons.

mRNA profiling gave opportunity to use transcripts of *HBD1* (human beta-defensin 1), *MUC4* (Mucin 4), *ESR1* (estrogen receptor 1), *MMP7*, *MMP10*, *MMP11* (matrix metalloproteinase 7, 10, 11) [2–8], and 16S–23S rRNA intergenic spacer regions of

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_3,

L. crispatus and *L. gasseri/L. johnsonii* as markers of vaginal fluid and menstrual blood [9–12]. None of the proposed individual markers proved to be fully specific to vaginal secretions nor present in all women. Therefore it was appropriate to combine the detection of bacterial and human tissue-specific transcripts in order to identify the vaginal fluid with higher reliability and specificity. On the basis of this assumption and previous reports [2, 3, 12, 13], we have developed a hexaplex PCR that is able to detect the following mRNAs: *HBD1*, *MUC4*, menstrual blood marker *MMP11*, housekeeping gene *G6PDH* (glucose 6-phosphate dehydrogenase), and the 16S–23S rRNA intergenic spacer regions (ISR) of *L. crispatus* and *L. gasseri/L. johnsonii* (Fig. 1).

The high variability of *HBD1*, *MUC4*, and *MMP11* expression, the changes in bacterial flora, and the controversy over the specificity of these markers do not allow for their sole use as a determinant of the presence of vaginal secretions. However, the simultaneous co-amplification of *HBD1*, *MUC4*, *MMP11*, 16S–23S rRNA of *L. crispatus* and *L. gasseri/L. johnsonii* and an endogenous control increases the credibility of the test.

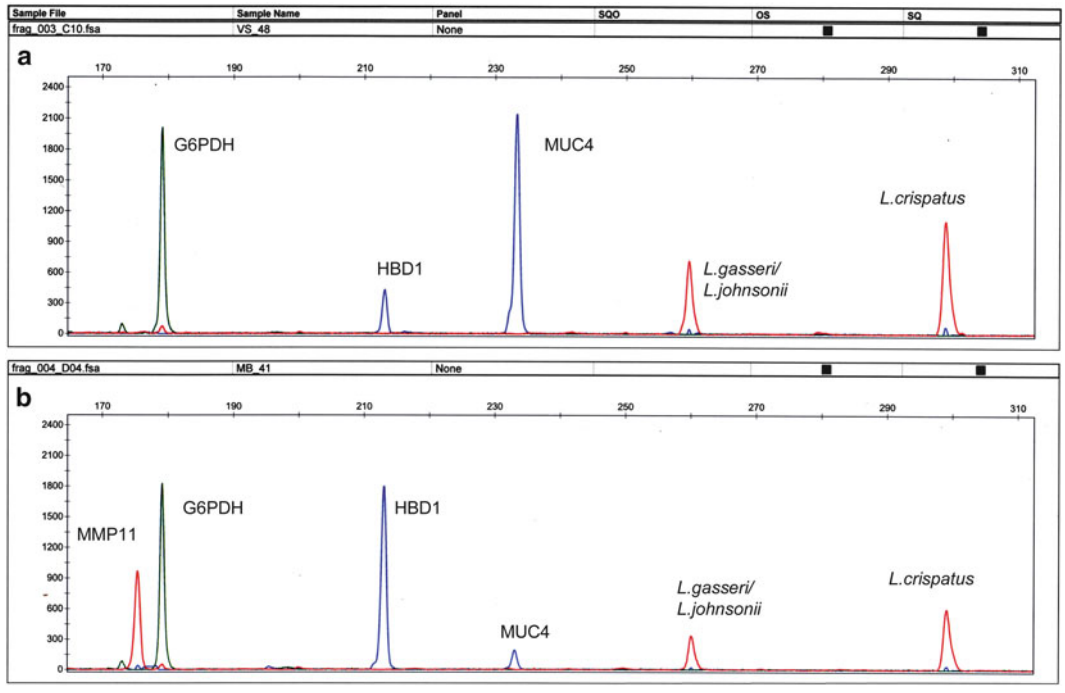


Fig. 1 Example of hexaplex assays for vaginal fluid and menstrual blood biomarkers detection (MMP11, HBD1, MUC4, L. gasseri/L. johnsonii, L. crispatus, and G6PDH). Panel A: A vaginal swab free of menstrual blood (22nd day of menstrual cycle). *Panel B*: A menstrual blood swab (2nd day of menstrual cycle)

2 Materials

2.1 General Supplies and Equipment

1. Vortex mixer.
2. Fume hood.
3. Microcentrifuge.
4. Thermocycler.
5. 310, 3100, 3130, or 3500 Genetic Analyzer (Life Technologies).
6. RNaseZap® wipes and spray (Ambion®).
7. Microcentrifuge tubes.
8. Spin baskets.
9. Forceps.

2.2 RNA Isolation

1. RNeasy Mini kit (Qiagen) (*see* **Note 1**). The required buffers are supplied with kit.

   a) 96–100 % ethanol must be added as per the manufacturer's instructions to the RPE buffer prior to use.
   b) 14.3 M β-mercaptoethanol (βME) is required for addition to RLT Plus buffer. A working solution of this buffer contains 10 μl βME per 1 ml of RLT Plus buffer. Add βME to an aliquot of the RLT Plus buffer rather than the entire bottle as the RLT Plus buffer/βME mixture is only stable at room temperature for 1 month.
   c) 70 % ethanol (dilute ethanol in RNase-free H_2O).
   d) RNase-Free DNase Set (Qiagen). If using RNase-Free DNase set for the first time, prepare DNase stock solution: using an RNase-free needle and syringe inject 550 μl RNase-free water into the DNase I vial (1500 Kunitz units) and mix gently. Do not vortex.

2.3 Reverse Transcription

1. SuperScript VILO cDNA Synthesis kit (Life Technologies) (*see* **Note 1**). All reagents are provided with the kit.

2.4 PCR

1. Fluorescently labeled primers (Table 1).
2. Multiplex PCR kit (Qiagen) (*see* **Note 1**).

2.5 Electrophoresis

1. Capillary electrophoresis platform, e.g., ABI PRISM 310, 3100, 3130, or 3500 Genetic Analyzers.
2. Hi-Di formamide (Life Technologies).
3. GeneScan 500 LIZ Size Standard (Life Technologies).
4. Performance-optimized polymer 4 (Life Technologies).

Table 1
Primers used in the hexaplex PCR reaction

	Primers sequences	Concentration of primers in PCR reaction	Expected amplicon length (bp)	Observed amplicon length (bp)	Reference
MMP11 Matrix metalloproteinase 11	F: 5′-PET-CAA GAC TCA CCG AGA AGG GG R: 5′-TAG CGA AAG GTG TAG AAG GCG	0.2 μM	17 4	175 (red)	[13]
HBD1 Human beta-defensin 1	F: 5′-6-FAM-CCC AGT TCC TGA AAT CCT GA R: 5′-CAG GTG CCT TGA ATT TTG GT	0.3 μM	215	213 (blue)	[2]
MUC4 Mucin 4	F: 5′-6-FAM-GGA CCA CAT TTT ATC AGG AA R: 5′-TAG AGA AAC AGG GCA TAG GA	0.02 μM	235	233 (blue)	[3]
L. gasseri/L. johnsonii 16S–23S rRNA ISR	F:5′-PET-GAG AAA GCC AAG CGG AAGC R:5′-TTG CTT ACT TAC TGC TCC CCG	0.1 μM	260–267	259–268 (*see* **Note 2**) (red)	[12]
L. crispatus 16S–23S rRNA ISR	F:5′-PET-CAG AGC AAG CGG AAG CAC A R:5′-CAT CTC TGC ATT GGG TTC CC	0.05 μM	299	299 (red)	[12]
G6PDH Glucose 6-phosphate dehydrogenase	F: 5′-VIC-ATC ATC GTG GAG AAG CCC TTC R: 5′-GTT CCA GAT GGG GCC GA	0.1 μM	180	179 (green)	[13]

3 Methods

3.1 RNA Isolation from Biological Stains

A significant problem in the course of RNA isolation is its low stability resulting from the presence of ribonucleases (RNases) that are difficult to inactivate. For this reason only RNase-free equipment and buffers can be used. Commercially available kits include RNase-free buffers and tubes, which greatly facilitate work. Otherwise buffers and vessels that come into contact with isolation must be subjected to diethylpyrocarbonate (DEPC), which inactivates the enzymes.

Although other methods can be used, we propose RNeasy® Mini kit with RNase-Free DNase Set for RNA isolation and SuperScript VILO kit for reverse transcription. If using RNeasy Mini Kit and RNase-Free DNase Set, follow the manufacturer's protocol (*see* **Note 3**).

1. Before starting, prepare DNase solution, add β-mercaptoethanol to RLT buffer and 70 % ethanol to RPE buffer as described in Subheading 2.1.
2. Cut a 1 × 1 cm piece of stain (on textile or swab) with sterile scissors and tweezers and transfer to a 2 ml microcentrifuge, RNase-free tube.
3. Add 400 μl RLT buffer (with βME) directly to the stain.
4. Vortex thoroughly for 10 min, add 400 μl ethanol, and transfer as much liquid as possible to RNase column.
5. Follow the RNeasy kit protocol with DNAse I digestion:
    - Centrifuge for 15 s at 10,000 × *g*, discard the flow-through.
    - Add 350 μl RW1 buffer to the column, centrifuge 15 s/10,000 × *g*, discard the flow-through.
    - Prepare DNase I solution (add 10 μl DNase I stock solution to 70 μl RDD buffer and mix gently).
    - Add 80 μl of DNase I solution (70 μl RDD plus 10 μl DNase I stock solution) to the column.
    - Add 350 μl RW1 buffer to the column, centrifuge 15 s/10,000 × *g*, and discard the flow-through.
    - Add 500 μl RPE buffer to the column, centrifuge 15 s/10,000 × *g*, and discard the flow-through.
    - Add 500 μl RPE buffer to the column, centrifuge 2 min/10,000 × *g*, and discard the flow-through.
    - Place RNeasy column in a new 1.5 ml tube. Add 30 μl RNase-free water and incubate at room temperature for 5 min.
    - Centrifuge for 1 min at 10,000 × *g*.

6. Measure RNA concentration using NanoDrop-1000 or other technique (*see* **Note 4**) and dilute it to 2 ng/μl. Immediately perform reverse transcription or store RNA extracts at −20 °C.

Perform all procedure on ice and work quickly to avoid RNA degradation.

3.2 Reverse Transcription

1. Prepare reaction mix using SuperScript VILO cDNA Synthesis Kit according to Table 2. Include a negative control (water), a positive control (vaginal RNA with a known, previously determined profile), and a reverse transcription control.
2. Add 18 μl reaction mix to 0.5 ml RNase-free microcentrifuge tubes and then add RNA.
3. Vortex gently and spin the samples.
4. Incubate the samples at 25 °C for 10 min, then at 42 °C for 60 min, finally inactivate at 85 °C for 5 min.
5. Store cDNA at −20 °C.

3.3 Multiplex PCR

1. Prepare primer mix according to Table 1 (*see* **Note 6**).
2. Prepare reaction mix according to Table 3. Include a negative control.
3. Add 18.5 μl of reaction mix into 0.2 ml microcentrifuge tubes.
4. Add 6.5 μl of cDNA and products of control reactions.
5. Perform PCR in the following conditions: initial denaturation at 95 °C for 15 min; 30 cycles of 94 °C/30 s, 55 °C/90 s, 72 °C/60 s; final elongation at 60 °C for 30 min.

3.4 Capillary Electrophoresis

1. Add 1 μl of PCR product to 12 μl of Hi-Di formamide and 0.5 μl GeneScan 500 LIZ Size Standard.

Table 2
Reverse transcription reaction components (see Note 5)

	Amount per sample	
	Sample/positive control/negative control	Reverse transcription control
5× VILO reaction mix	4 μl	4 μl
10× superscript enzyme mix	2 μl	–
RNA (2 ng/μl) or water	5 μl	5 μl
Nuclease-free water	9 μl	11 μl
Total reaction volume	20 μl	20 μl

Table 3
PCR reaction mix (see Note 5)

Reagent	Amount per sample
2× Qiagen multiplex PCR master mix	12.5 µl
Primers mix (Table 1)	6 µl
cDNA or nuclease-free water	6.5 µl
Total reaction volume	25 µl

2. Samples can be processed using ABI PRISM 310, 3100, 3130 and newer genetic analyzers. If using ABI PRISM 3130 machine the example conditions are: a default run module, Fragment Analysis36_POP4, and G5 dye set (POP-4 polymer, 5 s injection, 3.0 kV injection voltage, 60 °C, runtime 1000 s, filter set: G5). Analyze samples using GeneMapper or Genotyper software (Life Technologies).

3.5 Interpretation of the Results

Establish the analytical threshold for your test (*see* **Note 7**). Peaks below this value should not be considered as a positive result for particular markers. Check if the DNA size marker was called properly.

Table 1 shows product lengths for each pair of primers (*see* **Note 8**). Presence of the housekeeping gene *G6PDH* peak confirms human origin of biological material and proper amplification. If *G6PDH* product is not present, the test is invalid**.** The presence of peaks of at least three of the five markers and *G6PDH* leads to conclusion that the presence of vaginal fluid or/and menstrual blood (if *MMP11* is one of the positive markers) is very likely (*see* **Note 9**). *MMP11* is the marker of menstrual blood, which is present in vaginal secretion at the beginning of menstrual cycle. Due to the nature of the vaginal material, all the markers are considered to indicate the presence of vaginal fluid and menstrual blood. While a positive result for *MMP11* is a strong determinant of menstrual blood, a negative result should be treated as inconclusive when other markers are present (*see* **Note 10**).

4 Notes

1. In forensic genetics sensitivity of methods is of great importance because of the availability of only small amounts of biological material. Other methods than proposed can be used, provided they have similar or better sensitivity.
2. When using proposed primers, values from 259 to 268 bp should be considered as correct lengths of the *L. gasseri/L.*

johnsonii amplicon. However, due to possible differences in the bacterial strains present in different parts of the world, these amplicons length range should be treated carefully. The length of the *L. gasseri/L. johnsonii* amplicons can vary, even in the same woman during one menstrual cycle. Different strains with different amplicon lengths may coexist (Fig. 2) [14].

3. DNase treatment step is important and necessary although it may result in a decrease of RNA integrity. If DNA is not efficiently removed from RNA extract, artifacts and additional products on electropherograms could be observed. In proper protocol performance, when DNA is removed from RNA extract, *MMP11, HBD1, MUC4,* and *G6PDH* primers amplify only cDNA in the PCR reaction. In case of residual DNA in the RNA extract, *MMP11* and *G6PDH* primers will produce additional, longer amplicons of 317 and 357 bp respectively. *HBD1* and *MUC4* primers are specific only to cDNA. Remember that *MMP11*, *HBD1*, and *MUC4* genes are present in every nucleated human cell, and only PCR product unambiguously originating from cDNA amplification can be a true marker of vaginal material.

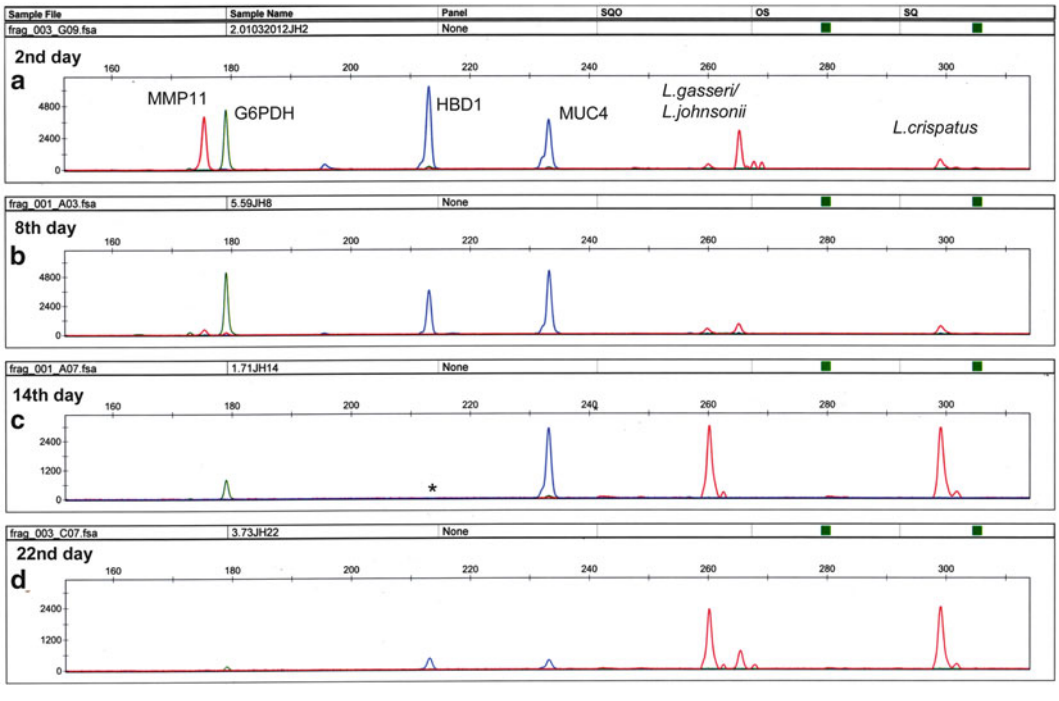


Fig. 2 The variable expression of vaginal fluid markers and bacterial flora during one menstrual cycle (2nd, 8th, 14th, and 22nd day of the cycle) of an exemplary woman. Variations in the length of the *L. gasseri/L. johnsonii* amplicon and changes in the expression of the markers, including different heights of the peaks are observed. Notice no HBD1 product on the 14th day of the cycle and no MMP11 product in the second half of the cycle

4. We have used NanoDrop-1000 to measure RNA concentration as it consumes only 1.7 μl of RNA sample. Although it is not the most accurate of available methods, in our experience it has given satisfactory results. Other technology may be used to obtain more reliable results. RNA integrity is not assessed.
5. Reaction was successfully performed at half volume.
6. Newly prepared primers mix should be stored at −20 °C in small volumes to avoid multiple freezing-thawing cycles.
7. In our case RFU of 50 units is used.
8. Be aware of artifacts like pull-ups or noisy base line that may appear and little differences between products length that may show up using different sequencers. Always check your primers after purchase.
9. We found large variations in the expression of the genes and bacterial flora among different women and during the menstrual cycle. Among the 41 vaginal swabs collected from different women in different stages of the menstrual cycle, *MMP11, HBD1, MUC4*, *L. gasseri/L. johnsonii*, *and L. crispatus* were detected in 36.6 %, 85.4 %, 95.1 %, 73.2 %, and 68.3 % of the samples, respectively. The endogenous control *G6PDH* was amplified in all cases. The analysis of swabs collected from five women (25–45 years old) on the 2nd, 8th, 14th, and 22nd day of the menstrual cycle did not show a clear correlation between the menstrual cycle and mRNA profile, except for the expression of *MMP11*, which only occurs up to the 8th day. As illustrated in Fig. 2, on the 2nd day of the cycle, all markers were present, while on the 14th day, there was no amplification of the *HBD1* transcript. The expression level of particular genes and the level of bacterial flora were changing, which is demonstrated by the variations in peak heights and the variability of lengths of *L. gasseri /L. johnsonii* amplicons [14].
10. The analysis of 18 menstrual blood swabs taken from four women on the first through fifth day of the menstrual cycle showed the presence of *MMP11* in 16 cases (89 %). In two cases, *MMP11* was not detected, although most or all of the other markers were amplified. In such a case, the presence of menstrual blood cannot be excluded, especially if a biochemical blood test is also positive [14].

Acknowledgments

This research was co-financed by the European Union through the European Social Fund within the framework of the Human Capital Operational Programme.

References

1. Virkler K, Lednev IK (2009) Analysis of body fluids for forensic purposes: from laboratory testing to non-destructive rapid confirmatory identification at a crime scene. Forensic Sci Int 188:1–17
2. Haas C, Klesser B, Maake C, Bär W, Kratzer A (2009) mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. Forensic Sci Int Genet 3:80–88
3. Juusola J, Ballantyne J (2005) Multiplex mRNA profiling for the identification of body fluids. Forensic Sci Int 152:1–12
4. Fang R, Manohar CF, Shulse C, Brevnov M, Wong A, Petrauskene OV et al (2006) Real-time PCR assays for the detection of tissue and body fluid specific mRNAs. Accessed 4 February 2014
5. Bauer M, Patzelt D (2008) Identification of menstrual blood by real time RT-PCR: technical improvements and the practical value of negative test results. Forensic Sci Int 174:54–58
6. Bauer M, Patzelt D (2002) Evaluation of mRNA markers for the identification of menstrual blood. J Forensic Sci 47:1278–1282
7. Juusola J, Ballantyne J (2007) mRNA profiling for body fluid identification by multiplex quantitative RT-PCR. J Forensic Sci 52:1252–1262
8. Setzer M, Juusola J, Ballantyne J (2008) Recovery and stability of RNA in vaginal swabs and blood, semen, and saliva stains. J Forensic Sci 53:296–305
9. Giampaoli S, Berti A, Valeriani F, Gianfranceschi G, Piccolella A, Buggiotti L et al (2012) Molecular identification of vaginal fluid by microbial signature. Forensic Sci Int Genet 6:559–564
10. Akutsu T, Motani H, Watanabe K, Iwase H, Sakurada K (2012) Detection of bacterial 16S ribosomal RNA genes for forensic identification of vaginal fluid. Leg Med 14:160–162
11. Benschop CCG, Quaak FCA, Boon ME, Sijen T, Kuiper I (2012) Vaginal microbial flora analysis by next generation sequencing and microarrays; can microbes indicate vaginal origin in a forensic context? Int J Legal Med 126:303–310
12. Fleming RI, Harbison S (2010) The use of bacteria for the identification of vaginal secretions. Forensic Sci Int Genet 4:311–315
13. Fleming RI, Harbison S (2010) The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids. Forensic Sci Int Genet 4:244–256
14. Jakubowska J, Maciejewska A, Pawlowski R, Bielawski KP (2013) mRNA profiling for vaginal fluid and menstrual blood identification. Forensic Sci Int Genet 7:272–278

Chapter 4

Preservation of and DNA Extraction from Muscle Tissue

Dennis McNevin

Abstract

As well as protecting DNA for subsequent analysis, tissue preservation methods ideally should be safe, readily available, and easy to transport at relatively low cost. Formalin (formaldehyde solution), used extensively to preserve medical and museum specimens, irreparably damages DNA. We have found four tissue preservatives (solid salt, salt-saturated dimethyl sulfoxide (DMSO)–EDTA solution, ethanol solution, and ethanol–EDTA solution) that preserved muscle tissue at 35 °C for up to 1 month: full short tandem repeat (STR) profiles were obtained after preservation. In addition, salt-saturated DMSO–EDTA solution yielded full STR profiles from aliquots of the liquid preservative surrounding muscle tissue.

Key words Tissue preservation, DNA extraction, DNA profiling, Salt, Ethanol, Dimethyl sulfoxide, DMSO, EDTA

1 Introduction

Tissue preservation for forensic purposes is most often associated with disaster victim identification with the intention of identifying and repatriating all victims and body parts. Forensic DNA analysis is one of the three primary methods of identification recommended by the International Criminal Police Organisation (INTERPOL), together with fingerprint and dental analysis [1], and can therefore play a crucial role for identifying victims. Mass disasters that occur in remote locations pose special problems [2–4] and those in tropical climates require samples to be processed quickly as degradation of bodies is accelerated.

Analysis should be as easy and efficient as possible and repeat analyzes due to problems like low template DNA and PCR inhibition should be avoided. Preservation methods should be able to generate a profile using commercial multiplex PCR chemistries. Skeletal muscle tissue is typically sampled for DNA analysis [1, 5] and requires preservation, from the time of collection to the point at which it can be transported to a laboratory refrigerator/freezer. INTERPOL guidelines state that formalin or formaldehyde

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_4,

solution should not be used as a preservative because it degrades DNA [1]. The International Society for Forensic Genetics also advises against the use of formalin [5].

The constituents of DNA preservatives should play a role in arresting DNA degradation (for a review, see [6]). NaCl is a common preservative that has been used for centuries. In solid form, it desiccates the sample that inactivates nucleases and slows microbial growth [7]. Other desiccants such as silica beads have a similar effect [7, 8]. When in an aqueous solution the NaCl also denatures proteins, including nucleases [7]. Chelating agents like EDTA bind to metal ions such as those required by nucleases for normal function (e.g., Mg^{2+} and Ca^{2+}) and therefore will slow or stop the nuclease activity on DNA [9, 10]. Detergents such as Tween 20 are also thought to lyse cells and inactivate nucleases, although the precise mechanism has not been described [11]. Ethanol removes water from the sample and denatures proteins and nucleases [9, 12]. Ethanol is also an antimicrobial agent and will protect against bacterial degradation [9]. Dimethyl sulfoxide (DMSO) is known to promote the dermal absorption of chemicals with which it is mixed [13, 14] so, while it does not directly preserve DNA, it enhances the absorption of other preservatives across biological membranes and into the cell [15]. Finally, buffering the pH stabilizes DNA by decreasing the rate of acid-catalyzed depurination [16, 17].

We were able to obtain full short tandem repeat (STR) profiles after preservation of muscle tissue at 35 °C for up to 1 month in four preservatives (solid salt, salt-saturated DMSO–EDTA solution, 70% ethanol solution, and 70% ethanol–EDTA solution) (Fig. 1 and *see* **Note 1**). In addition, DMSO solution yielded full STR profiles from aliquots of the liquid preservative surrounding muscle tissue (Fig. 2 and *see* **Note 2**). Full STR profiles (tissue and liquid preservative aliquots) were also obtained from two proprietary tissue preservatives: DNA Genotek Tissue Stabilising Kit (DNA Genotek, [18], *see* **Note 3**) and DNAgard (Biomatrica, [19]).

There are obvious advantages when able to extract DNA directly from a liquid preservative. The preserved tissue does not have to be handled, saving processing time and minimizing the opportunities for cross-contamination of samples. A micropipette can be used to withdraw an aliquot of preservative directly instead of withdrawing the tissue from the preservative and excising a portion for DNA extraction. A tissue sample brings with it associated cellular debris which can confound DNA extraction processes with "dirty" samples.

In support of our study, Michaud and Foran [20] found 70% ethanol and DMSO to be effective preservatives with greater than 90% amplification of a 257 bp target in porcine tissue. DMSO was significantly better at recovering longer DNA fragments (>90% amplification of a 642 bp target compared with <90% for ethanol).

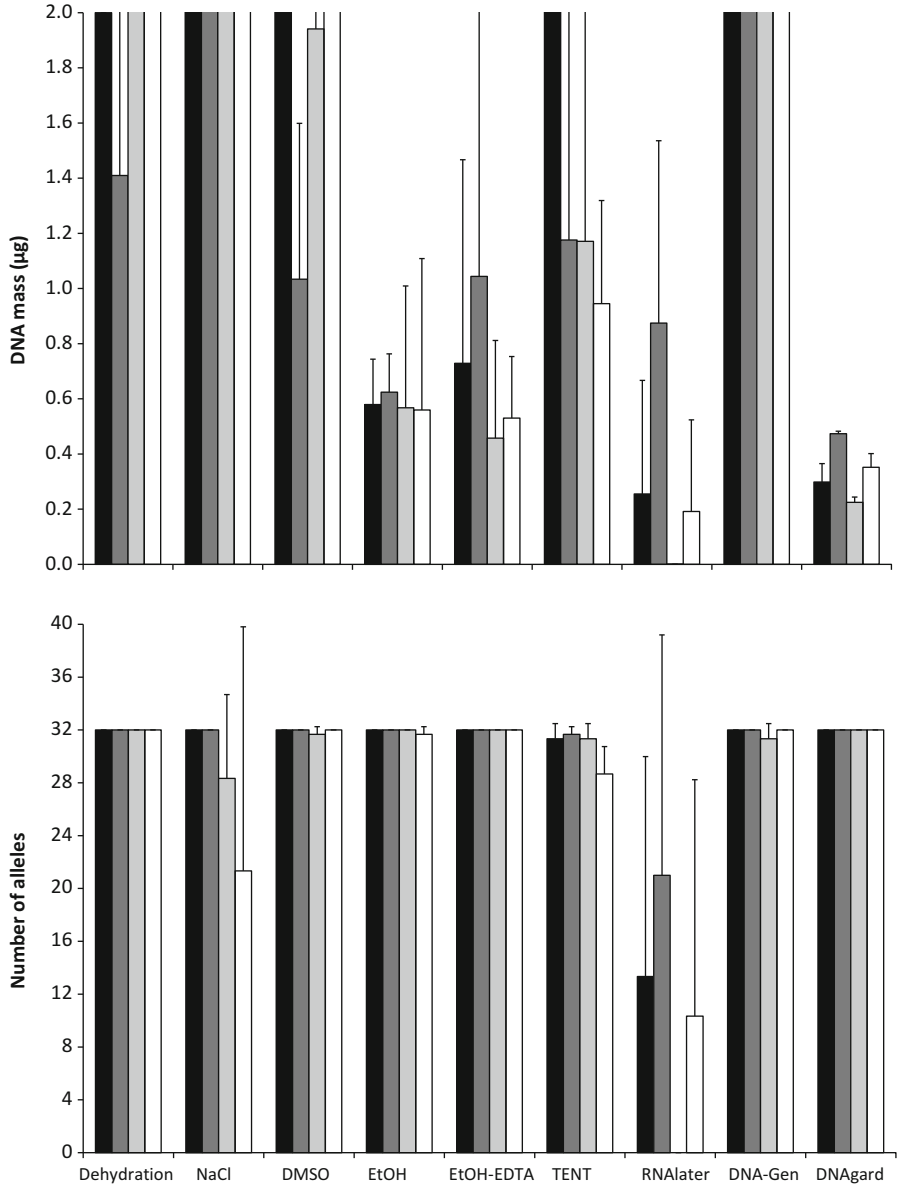


Fig. 1 Mean DNA masses extracted (*top*) and mean number of consensus alleles (maximum of 32, relative to tissue donors' buccal swabs) observed in genotypes (*bottom*) from excised tissue in different preservatives. Error bars represent one standard deviation. Reproduced from Allen-Hall and McNevin [24] with permission from Elsevier

While DMSO itself has a low toxicity [21], it readily penetrates the skin and facilitates the absorption of associated substances. Large quantities of preservative with this constituent present handling dangers although aqueous DMSO solution is nonflammable [21]. Ethanol is definitely flammable. Both these compounds may have restrictions placed on them when transporting to the scene of a

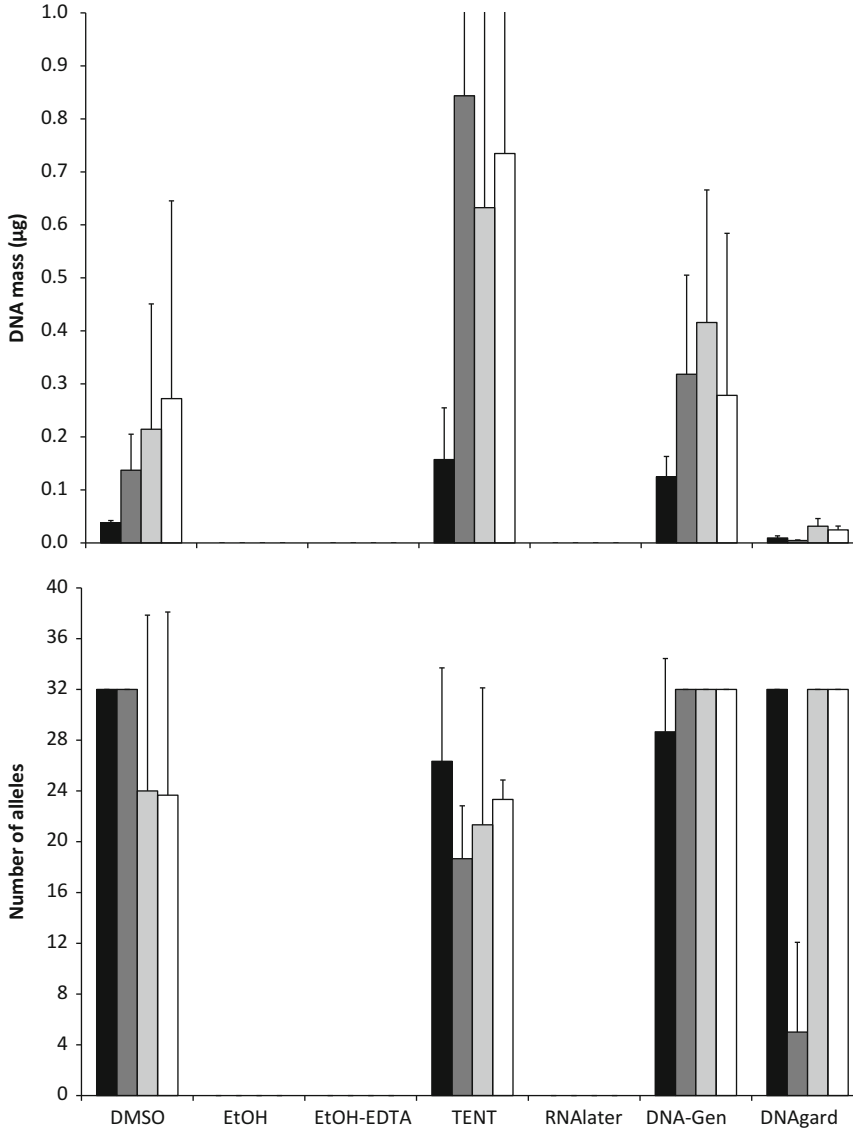


Fig. 2 Mean DNA masses extracted (*top*) and mean number of consensus alleles (maximum of 32, relative to tissue donors' buccal swabs) observed in genotypes (*bottom*) from 50 µl aliquots of preservatives surrounding tissue slices. Error bars represent one standard deviation. Reproduced from Allen-Hall and McNevin [24] with permission from Elsevier

disaster. Common salt, on the other hand, is widely available and poses no safety concerns with very little chance of spillage or evaporation over time.

We employed a standard phenol–chloroform/ethanol precipitation method to extract DNA from preserved tissues. In our hands, these extracts were considerably cloudier when derived from tissue samples than when derived from liquid preservative.

Other DNA extraction methods including most magnetic bead and silica column technologies are also appropriate. Most of these have their own lysis buffers with proteinase K and dithiothreitol (DTT) (*see* **Note 4**).

We have recently found that aliquots of preservative solution can be added directly to PCR, without the need for DNA extraction [22]. The DNA in the preservative is quantified using real-time PCR and then diluted appropriately for STR genotyping.

2 Materials

Solutions are prepared using purified water with a resistivity of at least 18 MΩ·cm at 25 °C. All reagents and chemicals used should be at least A.R. (Analytical Reagent) grade and of molecular biology grade wherever possible (the latter to reduce the risk of introducing nucleases which degrade DNA).

Reagents and microcentrifuge tubes should be autoclaved in advance, where appropriate, and handled in a laminar flow cabinet to minimize contamination with DNA from the operator. All surfaces should be cleaned with 0.4% sodium hypochlorite (bleach) and 70% ethanol and laminar flow cabinets should be irradiated with UV light for 30 min prior to use (*see* **Note 5**).

2.1 Preservatives

Percentage concentrations are volume percentages (v/v).

1. Salt: Laboratory grade NaCl (MW = 58.44 g/mol) (*see* **Note 6**).
2. DMSO solution: 20% DMSO, 0.25 M EDTA, saturated with NaCl, pH 8.0. Add 93 g disodium ethylenediaminetetraacetic acid dihydrate (Na_2·EDTA·$2H_2O$, MW = 372.24 g/mol) to 500 ml water and dissolve by heating and mixing (*see* **Notes 7** and **8**). Allow the EDTA solution to return to room temperature. Mix and adjust the pH to 8.0 using NaOH (*see* **Note 9**). Slowly add 200 ml dimethyl sulfoxide (DMSO, $(CH_3)_2SO$, MW = 78.13 g/mol) and 150 g salt (NaCl, MW = 58.44 g/mol) while mixing (*see* **Note 10**). Readjust the pH to 8.0 (*see* **Note 11**). Make up to 1 l with water (*see* **Note 12**). Gradually add NaCl while mixing until no more dissolves, while simultaneously adjusting the pH to 8.0 (*see* **Note 13**). Sterilize by autoclaving and store at room temperature. Allow any excess salt to settle before use.
3. Ethanol solution: 70% ethanol. Add 700 ml ethanol (CH_3CH_2OH, MW = 46.07 g/mol) to 300 ml water (*see* **Note 12**). Store at room temperature.
4. Ethanol–EDTA solution: 70% ethanol, 0.1 mM EDTA. Add 0.037 g of disodium ethylenediaminetetraacetic acid dihydrate

($Na_2{\cdot}EDTA{\cdot}2H_2O$, MW = 372.24 g/mol) to 300 ml water and stir until EDTA dissolves (*see* **Note 7**). Add 700 ml ethanol (CH_3CH_2OH, MW = 46.07 g/mol) (*see* **Note 12**). Store at room temperature.

2.2 DNA Extraction

1. TENT buffer: 10 mM Tris–HCl, 10 mM EDTA, 100 mM NaCl, 2% Tween 20, pH 8.0. Add 1.21 g of Tris base (MW = 121.14), 3.72 g disodium ethylenediaminetetraacetic acid dihydrate ($Na_2{\cdot}EDTA{\cdot}2H_2O$, MW = 372.24 g/mol), 5.84 g salt (NaCl, MW = 58.44 g/mol), and 20 ml Tween 20 to approximately 900 ml water (*see* **Note 7**). Mix and adjust the pH to 8.0 with NaOH (*see* **Note 9**). Make up to 1 l with water (*see* **Note 12**). Sterilize by autoclaving and store at room temperature.
2. TE buffer: 10 mM Tris–HCl, 0.1 mM EDTA, pH 8.0. Add 1.21 g of Tris base (MW = 121.14) and 0.037 g disodium ethylenediaminetetraacetic acid dihydrate ($Na_2{\cdot}EDTA{\cdot}2H_2O$, MW = 372.24 g/mol) to approximately 900 ml water (*see* **Note 7**). Mix and adjust the pH to 8.0 with HCl (*see* **Note 14**). Make up to 1 l with water (*see* **Note 12**). Sterilize by autoclaving and store at room temperature.
3. Sodium acetate: 3 M sodium acetate, pH 5.2. Add 24.61 g CH_3COONa (MW = 82.03 g/mol) to approximately 90 ml water. Mix and adjust pH to 5.2 with glacial acetic acid in a fume hood. Make up to 100 ml with water (*see* **Note 12**). Sterilize by autoclaving and store at room temperature (*see* **Note 15**).
4. Ethanol solution: 70% ethanol. Add 70 ml ethanol (CH_3CH_2OH, MW = 46.07 g/mol) to 30 ml water (*see* **Note 12**). Store at −20 °C.

3 Methods

Reagents and microcentrifuge tubes are autoclaved in advance, where appropriate, and handled in a laminar flow cabinet to minimize contamination with DNA from the operator. All surfaces are cleaned with 0.4% sodium hypochlorite (bleach) and 70% ethanol and laminar flow cabinets are irradiated with UV light for 30 min prior to use (*see* **Note 5**). Phenol/chloroform/isoamyl alcohol is handled in a fume hood to prevent exposure to toxic vapors.

3.1 Reference Samples

Reference samples should be collected where possible if they are required for identification and if they exist (*see* **Note 16**). These may be in the form of buccal (inside cheek epithelial cell) swabs, blood samples (preserved in heparin or EDTA), FTA cards [23], or other belongings likely to have accumulated DNA (e.g.,

toothbrushes). If these reference samples are not available from deceased victims, then they may be obtained from close genetic relatives.

3.2 Preservation of Tissue in the Field

1. Excise a sample of each tissue using sterile instruments. Ideally, deep muscle tissue, not exposed to the environment, is preferable.
2. Place approximately 300 mg of each tissue into sterile 10 ml screw top test tubes in either 4 g salt or 2 ml of liquid preservative (*see* **Note 17**). Ensure that the tissue is covered with salt or submerged in preservative (*see* **Note 18**). These tubes may then be stored and/or transported at ambient temperatures.

3.3 Preparation of Reference Samples and Controls

1. Employ negative controls consisting of empty microcentrifuge tubes and positive controls (e.g., 5 µl human female genomic DNA: Promega, G1521).
2. Remove reference buccal swab gauzes from their stems and place in individual, sterile 1.5 ml microcentrifuge tubes.
3. Punch a disk out of the sample area on FTA cards and place in individual, sterile 1.5 ml microcentrifuge tubes. FTA card punches can also be used directly in PCRs (see manufacturer protocol [23]).
4. Alternatively, place other reference items (e.g., tooth brush bristles) in individual, sterile 1.5 ml microcentrifuge tubes.

3.4 Preparation of Tissue Samples

In a sterile environment, excise a tissue slice from each preserved tissue sample and place in individual, sterile 1.5 ml microcentrifuge tubes (*see* **Note 19**).

3.5 Preparation of Salt-Saturated DMSO–EDTA Solution

Pipette 50 µl of the salt-saturated DMSO–EDTA preservative solution (surrounding each tissue sample) into individual, sterile 1.5 ml microcentrifuge tubes.

3.6 Extraction of DNA (See Note 20)

1. Lysis buffer is prepared *immediately prior* to use. Add 30 µl of 20 mg/ml proteinase K and 30 µl of 1 M DTT per ml of TENT buffer (*see* **Note 21**). Mix gently—do not vortex as proteinase K is sensitive to mechanical shear.
2. Add 400 µl of lysis buffer to each 1.5 ml microcentrifuge tube and mix by inverting. Incubate at 56 °C overnight with intermittent inversion.
3. The next day, add a further 10 µl of 20 mg/ml proteinase K and 10 µl of 1 M DTT to each tube and mix by inverting. Incubate at 56 °C for 1 h.
4. Add 400 µl of phenol/chloroform/isoamyl alcohol (24:25:1, pH 8.0) to each tube in a fume hood (*see* **Note 22**). Vortex

vigorously for 1 min and centrifuge at 12,000 × *g* for 15 min to separate the aqueous and solvent phases.

5. Remove 300 µl from the upper (aqueous) layer, taking care not to disturb the interface, and transfer to a new, sterile, 1.5 ml microcentrifuge tube (*see* **Note 23**).
6. For every 100 µl of aqueous layer removed, add 10 µl of sodium acetate (3 M, pH 5.2) and 300 µl of cold (–20 °C) absolute ethanol (*see* **Note 24**). Mix by inverting and incubate at –20 °C for 15 min.
7. Centrifuge at 12,000 × *g* for 45 min.
8. Decant the liquid immediately from each tube and add 60 µl of cold (–20 °C) 70% ethanol. Mix by inverting and incubate at –20 °C for 15 min.
9. Centrifuge at 12,000 × *g* for 15 min.
10. Decant the liquid immediately from each tube and air-dry the DNA pellet.
11. Resuspend each pellet in a 50 µl aliquot of TE buffer. Store at –4 °C if genotyping within 1 week or store at –20 °C if genotyping more than a week later.

3.7 Genotyping

Genotyping can be performed using any standard method. We typically quantitate the extracted DNA using Quantifiler™ Human DNA Quantitation Kit (Applied Biosystems) or equivalent. A DNA standard is used to prepare a dilution series to establish a standard curve and this is used to calculate DNA concentration for each sample. DNA extracts are then diluted appropriately to bring them within the recommended range for a genotyping kit (e.g., 0.05–0.125 ng/µl for AmpFlSTR® Identifiler®, Life Technologies). The amplified PCR product is analyzed by capillary electrophoresis (e.g., 3500 Genetic Analyzer, Life Technologies).

We have recently found that aliquots of preservative solution can be added directly to PCR, without the need for DNA extraction [22]. The DNA in the preservative is quantified using real-time PCR (e.g., Quantifiler) and then diluted appropriately for STR genotyping (e.g., Identifiler).

4 Notes

1. Complete AmpFlSTR® Identifiler® (Life Technologies) genotypes (32 consensus alleles, concordant with a reference genotype) were produced at 4, 7, 14, and 28 days in nearly all instances from 50 µl DNA extracts from three different tissues preserved in solid NaCl, salt-saturated DMSO–EDTA solution, 70% EtOH, and 70% EtOH–EDTA [24].

2. DMSO consistently produced complete genotypes from a 50 µl aliquot of the liquid preservative. The ethanol-based preservatives (70% EtOH and 70% EtOH–EDTA), while yielding DNA from tissue, failed to produce DNA in extracts from the liquid preservative [24].
3. Another preservative manufactured by DNA Genotek (GenoFix™), although not currently available, has been shown to preserve DNA for extended periods of time and allow full STR profiling [25].
4. We have also successfully extracted DNA using the following commercial kits: ChargeSwitch® Nucleic Acid Purification Technology (Life Technologies), QIAamp DNA Investigator Kit (QIAGEN), ISOLATE II Nucleic Acid Kits (BIOLINE).
5. We use a G30T8 30 W germicidal UV-C lamp (Sankyo Denki) in a Class II Biological Safety Cabinet (Gelaire).
6. Commercial salt could be used but there is a risk of DNA contamination from reagents that are not subject to appropriate quality assurance.
7. Many dry reagents are hygroscopic and will absorb moisture from the atmosphere while they are being weighed, thus making accurate weight determination problematic. To avoid this, use fully hydrated reagents wherever possible. Hence use EDTA·$2H_2O$ (ethylenediaminetetraacetic acid dihydrate) instead of EDTA. The degree of hydration will vary the molecular mass of EDTA, as will its complementary cation, that is, whether it is disodium EDTA (Na_2·EDTA) or tetrasodium EDTA (Na_4·EDTA). This will alter the mass of EDTA required for 0.25 M concentration.
8. High concentrations of EDTA will not dissolve in water at room temperature.
9. Adjust first with a lower concentration of NaOH (e.g., 1 M). If no appreciable change in pH occurs, adjust with a higher concentration (e.g., 5 M). As desired pH is approached, switch back to lower concentration for fine adjustment. This will avoid overshooting the desired pH (it can be readjusted downwards with HCl if required).
10. Addition of DMSO at acidic pH will precipitate EDTA.
11. Addition of DMSO and salt will alter the pH [26].
12. For other volumes, multiply all reagent amounts accordingly.
13. A further 50 g NaCl may be required.
14. Adjust first with a lower concentration of HCl (e.g., 0.1 M). If no appreciable change in pH occurs, adjust with a higher concentration (e.g., 1 M). As desired pH is approached, switch back to lower concentration for fine adjustment. This will avoid overshooting the desired pH.

15. Smaller aliquots can be frozen for long term storage.
16. When field collection is necessary, the conditions should be as controlled as possible, ideally in a mortuary, to avoid cross-contamination of samples.
17. We use polypropylene urine collection tubes (Sarstedt, Cat. # 62.9924.284).
18. The mass of tissue and the volume of preservative are not critical. What is important is that the tissue is submerged. For the salt-saturated DMSO–EDTA solution, as little preservative as is required to submerge the tissue will maximize the yield of DNA from the preservative solution.
19. 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 (approximately 10% of our total collected tissue).
20. We use a phenol–chloroform/ethanol precipitation method to extract DNA from preserved tissues. Other DNA extraction methods including most magnetic bead and silica column technologies are also appropriate. Most of these have their own lysis buffers with proteinase K and DTT.
21. Proteinase K and DTT solutions are available commercially but can also be prepared in the laboratory. Proteinase K solutions must contain calcium ions for activity (e.g., 1 mM calcium chloride).
22. Phenol/chloroform/isoamyl alcohol is hydrophobic and will drip from hydrophilic pipette tips.
23. The exact volume is not critical. What is important is that phenol/chloroform/isoamyl alcohol is not carried over with the aqueous solution removed. It is better to withdraw less solution than to risk carry over. If the aqueous solution removed is not clear, fresh phenol/chloroform/isoamyl alcohol can be added and the aqueous phase removed again.
24. Optional: Also add 10 μl of 125 mM EDTA (autoclaved) for every 100 μl of aqueous phase removed. It is important to keep the overall concentration of ethanol >65%.

Acknowledgments

The author gratefully acknowledges funding from the Australian Federal Police Forensic and Data Centres. Collection of skeletal muscle tissues was generously facilitated by Associate Professor Paul Smith and Dr. Jennie Scarvell from the Trauma and Orthopaedic Research Unit (TORU), Canberra Hospital, Australia, and Ms. Maria Hartley, ACT Bone Bank, Canberra, Australia.

References

1. INTERPOL (2009) Disaster victim identification guide. INTERPOL, Lyon
2. Morgan OW, Sribanditmongkol P, Perera C et al (2006) Mass fatality management following the South Asian tsunami disaster: case studies in Thailand, Indonesia and Sri Lanka. PLoS Med 3(6):808–815
3. Tsokos M, Lessig R, Grundmann C et al (2006) Experiences in tsunami victim identification. Int J Leg Med 120(3):185–187
4. Gilpin G (2009) Personal account from the ice—Inspector Gilpin. New Zealand Air Line Pilot's Association. http://www.erebus.co.nz/TheAccident/TheRecoveryOperation/InspectorGilpinsAccount.aspx. Accessed 16 Feb 2016
5. Prinz M, Carracedo A, Mayr WR et al (2007) 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 Suppl Series 1(1):3–12
6. Allen-Hall A, McNevin D (2013) Non-cryogenic forensic tissue preservation in the field: a review. Aust J Forensic Sci 45(4):450–460
7. Nagy ZT (2010) A hands-on overview of tissue preservation methods for molecular genetic analyses. Org Divers Evol 10(1):91–105
8. Grassberger M, Stein C, Hanslik S et al (2005) Evaluation of a novel tagging and tissue preservation system for potential use in forensic sample collection. Forensic Sci Int 151(2–3):233–237
9. Seutin G, White BN, Boag PT (1991) Preservatin of avian blood and tissue samples for DNA analysis. Can J Zool 69:82–90
10. Giannakis C, Forbes IJ, Zalewski PD (1991) Ca^{2+}/Mg^{2+}-dependent nuclease: tissue distribution, relationship to inter-nucleosomal DNA fragmentation and inhibition by Zn^{2+}. Biochem Biophys Res Commun 181(2):915–920
11. Muralidharan K, Wemmer C (1994) Transporting and storing field-collected specimens for DNA without refrigeration for subsequent DNA extraction and analysis. Biotechniques 17(3):420, 422
12. Flournoy LE, Adams RP, Pandy RN (1996) Interim and archival preservation of plant specimens in alcohols for DNA studies. Biotechniques 20(4):657–660
13. Stoughton RB, Fritsch W (1964) Influence of dimethylsulfoxide (DMSO) on human percutaneous absorption. Arch Dermatol 90:512
14. Elfbaum SG, Laden K (1968) The effect of dimethyl sulfoxide on percutaneous absorption: a mechanistic study, part I. J Soc Cosmet Chem 19:119–127
15. Kilpatrick CW (2002) Noncryogenic preservation of mammalian tissues for DNA extraction: an assessment of storage methods. Biochem Genet 40(1–2):53–62
16. Suzuki T, Ohsumi S, Makino K (1994) Mechanistic studies on depurination and apurinic site chain breakage in oligodeoxyribonucleotides. Nucleic Acids Res 22(23):4997–5003
17. Lindahl T, Nyberg B (1972) Rate of depurination of native deoxyribonucleic acid. Biochemistry 11(19):3610–3618
18. DNA Genotek (2016) Superior samples—proven performance. DNA Genotek, Inc. http://www.dnagenotek.com/ROW/index.html. Accessed 16 Feb 2016
19. Biomatrica (2016) DNAgard® tissue. Biomatrica. http://www.biomatrica.com/dnagardtissue.php. Accessed 16 Feb 2016
20. Michaud CL, Foran DR (2011) Simplified field preservation of tissues for subsequent DNA analyses. J Forensic Sci 56(4):846–852
21. Vignes R (2000) Dimethyl sulfoxide (DMSO): a "new" clean, unique, superior solvent. American Chemical Society, Washington, DC
22. Sorensen A, Berry C, Bruce D et al (2015) Direct-to-PCR tissue preservation for DNA profiling. Int J Leg Med. doi:10.1007/s00414-015-1286-z
23. Whatman Ltd (2016) FTA elute sample collection cards. GE Healthcare Life Sci. http://www.gelifesciences.com/webapp/wcs/stores/servlet/catalog/en/GELifeSciences-au/products/AlternativeProductStructure_18503. Accessed 16 Feb 2016
24. Allen-Hall A, McNevin D (2012) Human tissue preservation for disaster victim identification (DVI) in tropical climates. Forensic Sci Int Genet 6(5):653–657
25. Fregeau CJ, Vanstone H, Borys S et al (2001) AmpFlSTR profiler plus and AmpFlSTR COfiler analysis of tissues stored in GenoFix, a new tissue preservation solution for mass disaster DNA identification. J Forensic Sci 46(5):1180–1190
26. Matthews WS, Bares JE, Bartmess JE et al (1975) Equilibrium acidities of carbon acids. VI. Establishment of an absolute scale of acidities in dimethyl sulfoxide solution. J Am Chem Soc 97(24):7006–7014

Chapter 5

DNA Extraction: Organic and Solid-Phase

Wafa Altayari

Abstract

DNA extraction remains a critical step in DNA profiling of biological material recovered from scenes of crime. In the forensic community several methods have gained popularity, including Chelex®, organic extraction, and solid-phase extraction. While some laboratories streamlined their processes and only use one method we have retained several methods and continue to use these for different sample types. In this chapter we present three methods that have been used for several years in our laboratory.

Key words DNA extraction, Phenol–chloroform, Solid-phase extraction, PrepFiler®, EZ1® DNA Investigator®

1 Introduction

After the detection/characterization and collection of biological material the next step is the extraction of DNA, sometimes along with the extraction of RNA (*see* Chapters 2 and 3). In the field of molecular biology there are a multitude of methods available for DNA extraction, many of which are variations on the same principles. Chelex® 100 was one of the first methods to be adopted by the forensic community in the PCR era [1]. It is simple, inexpensive, and yields relatively high amounts of DNA [2]. However, it does have a limited capacity to remove PCR inhibitors [3]. Organic extraction, in particular, phenol–chloroform-based extraction, had widespread popularity in molecular biology, and yields high-molecular weight DNA (as long as the starting material is not degraded); when processing forensic samples the organic methods have proven to be effective at removing many PCR inhibitors (*see* Chapters 6 and 7). The popularity of organic methods has reduced in recent years, largely due to health and safety concerns because phenol is highly toxic, and also because of the limited potential for automation. Solid-phase DNA extraction techniques have become increasingly popular in forensic laboratories; these methods benefit from the potential for automation and also being efficient at removing

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_5,

many PCR inhibitors [3–8]. The choice of method to be employed for DNA extraction depends on several variables, including the likely amount of DNA in the samples, the substrate that the sample is on, the type of biological material, and the presence of potential PCR inhibitors [3, 9, 10].

2 Organic Extraction (Phenol–Chloroform)

Despite the toxicity of phenol and the extra burden that this places onto laboratories in terms of specialist equipment, safety precautions and disposal of spent reagents the phenol–chloroform method is still popular in many forensic laboratories especially when extracting DNA from poor quality samples. The sample types that we have used this method for are detailed in Table 1.

2.1 Materials and Equipment

1. DNA-free 1.5 ml tubes.
2. Phase-lock 2 ml tubes (*see* **Note 1**).
3. DNA-free scalpel/scissors.
4. DNA extraction buffer (2 % SDS, 0.01 M EDTA pH 8.0).
5. Proteinase K, 20 mg/ml (Sigma-Aldrich).
6. Dithiothreitol (DTT) (0.4 M) (Sigma-Aldrich).
7. Phenol–chloroform–isoamyl alcohol (25:24:1) (Sigma-Aldrich).
8. *n*-Butanol (Sigma-Aldrich).
9. Microcon® 100 columns (Merck Millipore).
10. Benchtop vortex.
11. Benchtop minifuge.
12. Incubator (37–56 °C).

3 Method

1. Prepare the sample according to Table 1.
2. In the sample tube add the required amount of extraction buffer, typically we use 400 µl, but the reaction volume can be scaled up if necessary.
3. Add Proteinase K enzyme (20 mg/ml) in a ratio of 10:1 (extraction buffer: Proteinase K), for example, when using 400 µl of extraction buffer then 40 µl Proteinase K should be added.
4. Add the reducing agent dithiothreitol (0.4 M DTT) if semen, hair, nails, or bone is present in the extraction—add the same volume as proteinase K.

Table 1
Sample preparation using the organic extraction protocol developed by Abu Dhabi Forensic Evidence Department

Sample type	Sample preparation
Whole blood	Thaw and equilibrate up the whole blood sample at room temperature (15–25 °C). Transfer 10 µl of the sample to a 1.5 ml tube.
Buccal swabs	Cut 1/4 - 1/2 of the swab and place it into a 1.5 ml tube.
Blood swabs	If the blood on the swab is diluted the entire swab will be cut and placed in 1.5 ml tube. If the blood swab is concentrated, then a small part from the swab will be cut and placed in 1.5 ml tube
Saliva on fabric	Cut out an area indicated for saliva (25 mm^2).
Blood on fabric	Cut out an area indicated for blood (25 mm^2).
Manufactured cigarettes	Place 0.5–1 cm paper from the end of the filter in 1.5 ml tube.
Roll-ups cigarettes (smoked)	Place 1 cm of the paper taken around the base of the cigarette in 1.5 ml tube. Avoid taking any of tobacco as this will inhibit the extraction and PCR.
Roll-ups cigarettes (unsmoked)	Take 1–2 cm of the seam. Separate it before putting it in 1.5 ml tube.
Stamps and envelops	Access the gum from the reverse of the baking paper by making 4 dots at the corner of the stamp that you can see through the reverse backing paper. Use a sterilize scalpel and very gently score between the dots without going through the other side. Carefully peel the backing paper off to reveal the gum layer; the surface texture should go from fibrous paper to shiny, smooth and slightly tinted. Carefully peel away the gum layer till leaving the stamp intact. Put this paper in 1.5 ml tube.
Chewing gum	Cut third to half of the gum and then place it in 1.5 ml tube. Smear the gum around the inside of the tube using a DNA-free metal rod.
Hair	Place 1 cm from the root end in the 1.5 ml tube.
Fingernail scrapings/ cuttings	In case of scrapings, do not take more than will half fill 1.5 ml tube. In case of cutting, chop the cuttings into small pieces and this place it in 1.5 ml tube. Don't take more than will half fill a 1.5 ml tube.
Bone\teeth\muscle	Add 50 mg of powdered bone\tooth into 1.5 ml tube. Add small pieces of muscle into 1.5 ml tube

5. Vortex the sample for 5 s and then centrifuge briefly.
6. Incubate the sample either at 37 °C overnight or at 56 °C for 2 h.
7. Before sample purification with phenol–chloroform filter any samples that contain absorbent materials using the Spinaroo process, e.g., swabs or fabrics (*see* **Note 2**).

8. Add phenol–chloroform–isoamyl alcohol (25:24:1) to the sample tubes with the equivalent amount of the extraction buffer with the Proteinase K and DTT. Vortex and centrifuge at 14,000 × *g* for 5 min. Repeat this step 2–3 times until no pellicle is visible in the interphase (*see* **Note 1**).
9. Add an amount of *n*-butanol equal to the amount of the phenol–chloroform–isoamyl alcohol (*see* **Note 3**).
10. Vortex and then centrifuge at 14,000 × *g* for 5 min.
11. Transfer the lower aqueous phase to a Microcon® 100 Centrifugal Filter to re-concentrate the DNA. Centrifuge at 4000 × *g* for 20–30 min, or until the liquid has been reduced to the minimum retained volume (*see* **Note 4**).
12. Wash the filtrated Microcon with 500 μl of DNA-free water and centrifuge at 4000 × *g* for 20–30 min. Repeat this **steps 2–3** times.
13. Elute the DNA in 25–50 μl of DNA-free water (TE (Tris–EDTA) buffer could also be used) by inverting the Microcon into a 1.5 ml tube. Centrifuge at 4000 × *g* for 3–5 min.

The scheme for DNA extraction from epithelial cells using organic extraction is shown in Fig. 1.

4 Solid-Phase Nucleic Acid Extraction

Solid-phase nucleic acid purification is now widely used as the cleanest and easiest technique to separate DNA from inhibitors. Many commercial kits are available that use solid-phase extraction. It has gained popularity based on the increased levels of automation that are possible, the ability of the methods to remove many types of PCR inhibitor and the relatively high yields of DNA that can be recovered from many types of evidence. The methods all rely on DNA being released from the cellular material, and then binding to a solid substrate; for example, magnetic silica-based beads with QIAamp chemistry and polymer embedded magnetic beads with PrepFiler® chemistry (Life Technologies). After binding non-DNA components can be washed away, before releasing the DNA from the solid support to yield high purity DNA. The technology is suitable for a wide variety of samples (Table 2).

4.1 Qiagen Investigator® Kit Using the EZ1® Bio-Robot Workstation

Qiagen corporation has developed an automate method for DNA extraction based on solid-phase extraction using silica coated beads [11].

4.2 Materials and Equipment

1. DNA-free 1.5 ml tubes.
2. G2 Lysis Buffer (provided with the kit).

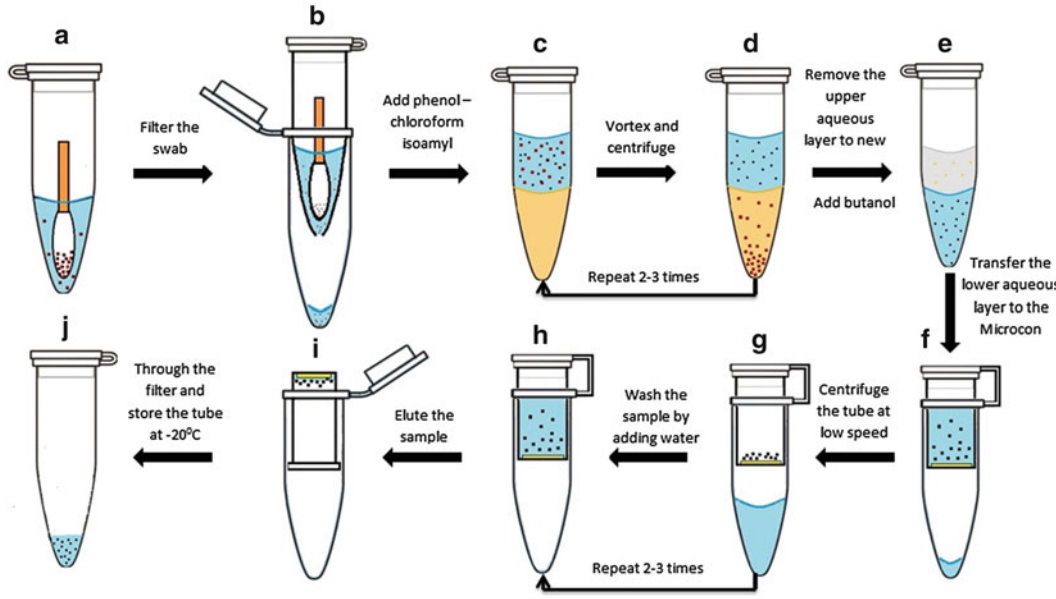


Fig. 1 Organic-based DNA extraction from epithelial cells (swab) using phenol–chloroform. (**a**) Cellular material is added to the lysis buffer and proteinase K and incubated at 56 °C for at least 2 h; (**b**) The digested swab is filtered by making a hole at the bottom of the tube and then placed it in a new 1.5 ml tube (spinaroo). The tube assembly is centrifuged at 4000 × *g*; (**c**–**d**) Phenol–chloroform–isoamyl alcohol is added to the filtrate solution, the solution is then vortexed and centrifuged. This step is repeated 2–3 times until there is no visible material at the interface (phase-lock gel tubes can be used when carrying out this step); (**e**) *n*-Butanol is added; this removes any residual of phenol–chloroform–isoamyl alcohol, which would inhibit downstream process such as PCR; (**f**) The aqueous phase is transferred to the Microcon 100 Centrifugal Filter assembly and centrifuged at low speed. (**g** and **h**) Sample is then washed using DNA-free water and centrifuged at 4000 × *g* for 20–30 min. This step is repeated 2–3 times depending on the sample condition and purity. Elution buffer is added to the Microcon and then the Microcon inverted in the elution tube. The inverted Microcon assembly is centrifuged for 3–5 min to collect the DNA at the bottom of the elution tube. (**i**) Elution tube is closed and stored at 4 °C short-term until the downstream analysis is finished. After that the samples are stored at –20 °C long-term

3. Proteinase K (provided with the kit).
4. Ultrapure (DNA-free) water.
5. EZ1 sample tubes, elution tubes, tips and holders, and cartridge.
6. Benchtop vortex.
7. Benchtop minifuge.
8. Heat block (56–95 °C).
9. EZ1 Bio-Robot workstation.

Table 2
Sample types suitable for processing as recommended by the manufacturers and validated for use under ISO17025 in Abu Dhabi Police Laboratories

	Manufacturers' recommendations		Internally validated	
Sample type	**PrepFiler®**	**DNA Investigator Kit**	**PrepFiler®**	**DNA Investigator Kit**
Whole blood	√	√	NT*	√
Dried blood	√	√	√	√
Buccal cells	√	√	NT	√
Surface and contact cells		√	√	X
Nails scraping		√	√	X
Chewing gum	√	√	√	√
Cigarette butts	√	√	√	√
Postage stamps and envelopes		√	√	NT
Hair	√	√	√	X
Bones or teeth	√	√	√	X
Soil		√	NT	NT
Stains on fabric	√	√	√	√
Human tissue		√	NT	NT
Epithelial cells mixed with sperm cells	√	√	√	X
Tape lifts	√		NT	NT

* NT stands for "Not Tested".

5 Method

1. Prepare the sample according to Table 3.
2. Pre-digest the samples by adding the recommended amount of lysis buffer and proteinase K provided with the kit (*see* Table 4). Dilute the lysis buffer with ultrapure water 1:1 when extracting DNA from absorbent material (*see* **Note 6**).
3. Vortex the samples thoroughly for 10 s and pulse spin them.
4. Incubate the sample under the recommended temperature and lysis time (Table 4).
5. Vortex the sample tube once or twice during the incubation or place it in the shaking incubator.
6. Incubate the sample at 95 °C for 5 min if you want to increase the yield of DNA.

Table 3
Shows sample preparation for samples to be used with EZ1® DNA Investigator® Kit. Adapted from manufacturer's manual—additional sample types are detailed in the manufacturer's manual)

Sample type	Sample preparation
Whole blood	Thaw and equilibrate up the whole blood sample at room temperature (15–25 °C). Transfer 10 µl of the blood sample to a 2 ml EZ1 sample tube, add 10 µl Proteinase K and 180 µl G2 Buffer (*see* **Note 5**).
Dried blood on FTA cards or Fabric	Punch 4 disks from the dried blood sample, each one is 3 mm in diameter and place it in 1.5 ml tube. If blood is on fabrics cut approximately 25 mm^2 section.
Buccal Swabs	Place 1/2 – 1/4 of the buccal swab into a 1.5 ml tube.
Chewing gum	Cut third to half of the gum and then place it in 1.5 ml tube. Smear the gum around the inside of the tube using a DNA-free metal rod.
Cigarette butts	Place 1 cm^2 paper from the end of the filter into 1.5 ml tube.
Stains on fabric	Place the fabric sample (25 mm^2) into 1.5 ml tube.

Table 4
Recommended protocols when using the EZ1 DNA Investigator Kit. Adapted from the manufacturer's manual—some variations from the manufacturer's recommendations are noted

Sample type	Recommended amount of lysis buffer (G2)	Recommended amount of Proteinase K (µl)	Recommended incubation time and temperature	Recommended EZ1 protocol
Whole blood (*see* **Note 5**)	180 µl Buffer G2	10	At 56 °C for 15 min	Trace protocol
Dried blood on FTA or fabric	190 µl Diluted buffer G2	10	At 56 °C for 15 min	Trace protocol (*see* **Note 7**)
Buccal swabs	190 µl Diluted buffer G2	10	At 56 °C for 15 min	Trace protocol (*see* **Note 7**)
Chewing gum	190 µl Buffer G2	10	At 56 °C for 15 min	Trace protocol
Cigarette butts	190 µl Diluted buffer G2	10	At 56 °C for 15 min	Trace protocol (*see* **Note 7**)
Stains on fabric	190 µl Diluted buffer G2	10	At 56 °C for 15 min	Trace protocol (*see* **Note 7**)

7. Flick or centrifuge the tube briefly to remove drops from inside the lid.
8. When extracting DNA from absorbent materials and using the trace protocol transfer the supernatant from the predigested tube (1.5 ml tube) to a 2 ml EZ1 tube using spinaroo method (*see* **Note 2**).

9. Load 2 ml EZ1 sample tubes, elution tubes, tips and holders, and cartridge into EZ1 Bio-Robot according to manufacturer's instructions.
10. Select the EZ1 Trace protocol.
11. Start the instrument according to manufacturer's instructions.
12. In the instrument:

    First, the sample will be added to the lysis buffer in the cartridge. This lysis buffer contains: guanidine thiocyanate (GuCN) and guanidine hydrochloride (GuCl) which work as chaotropic agents that lysis any remaining cells, denature the proteins, inhibit nuclease enzymes as well as provide the environmental conditions to promote the binding of the DNA to the paramagnetic silica beads. Secondly, the paramagnetic beads will be added to the lysis buffer. After that, the bound DNA will be washed to remove impurities twice and finally eluted under an alkaline condition using Tris–EDTA buffer. The extraction using the EZ1 instrument will take approximately 20 min. A schematic example of DNA extraction from epithelial cells using EZ1 instrument is shown below (Fig. 2).

6 PrepFiler® Express™ and Express BTA™ Forensic DNA Extraction Kits Using AutoMate Express™

PrepFiler® was initially introduced as a manual DNA extraction method; subsequently Applied Biosystems developed the AutoMate Express™, a small liquid handling robot similar to Qiagen EZ1 robot. This instrument use PrepFiler® chemistry which is similar to the EZ1 Qiagen kits except for the binding material. PrepFiler® uses encapsulated magnetic beads with polymer. This type of beads creates a large surface area with higher DNA binding capacity, which improved the extraction yield from some challenging samples. We use two kits: the PrepFiler® Express™ and PrepFiler® Express BTA™ Forensic DNA Extraction Kits.

6.1 Materials and Equipment

1. DNA-free 1.5 ml tubes.
2. PrepFiler lysis solution (provided with the kit).
3. Proteinase K (provided with the kit).
4. 1.0 M DTT (Sigma-Aldrich).
5. PrepFiler Lysep, PrepFiler sample tube, Bone/Tooth Lysate tube, extraction cartridge.
6. Benchtop vortex.
7. Benchtop minifuge.
8. Shaking Incubator (56–95 °C).
9. AutoMate Express™.

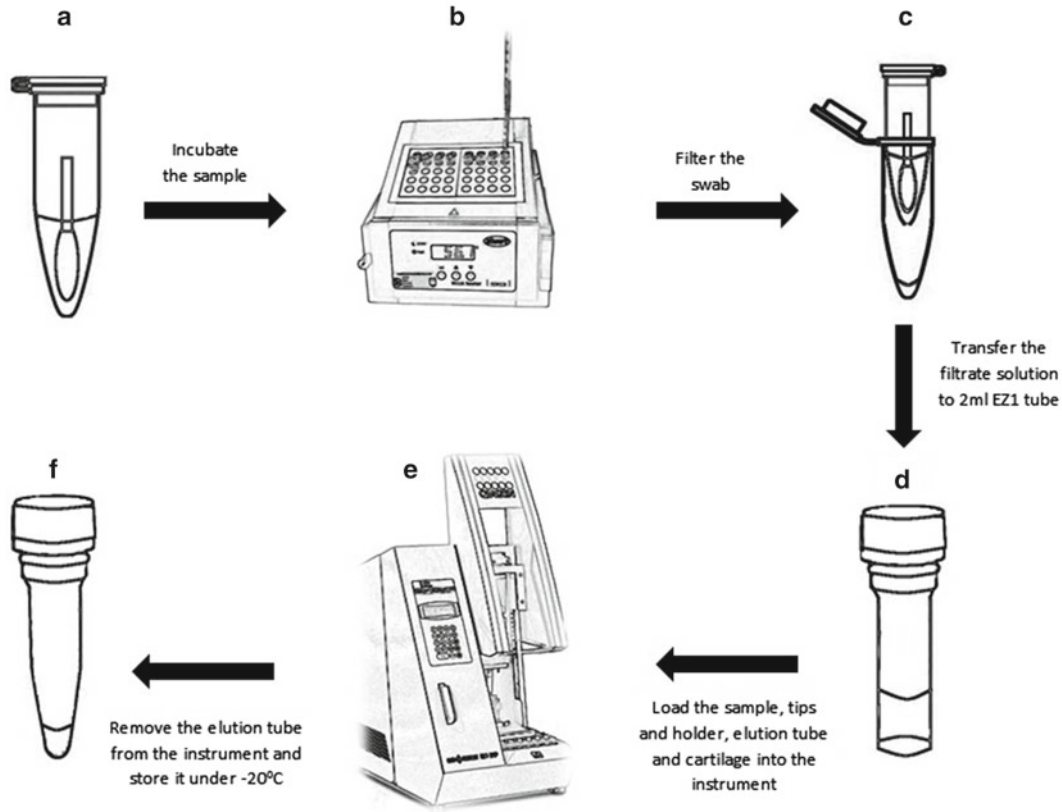


Fig. 2 DNA extraction from epithelial cells (swab) using solid-phase extraction method based on paramagnetic silica beads. (**a**) Cellular material is added to the lysis buffer and proteinase K which provided with the Qiagen kit. (**b**) The sample is then incubated at 56 °C for 15 min. (**c**) The digested swab is filtered by making a hole at the bottom of the tube and then placed it in a new 1.5 ml tube. Spin columns can also be used in this step. (**d**) Filtrate solution is transferred to 2 ml EZ1 tube. (**e**) The pre-digested sample, elution tube, tips and holders, and cartridge are loaded to the EZ1 instrument. The extraction process started using trace protocol installed in the instrument. The Tip Dance protocol can be used if the filtration step was not carried out. (**f**) After the run is complete the elution tube is closed and stored at 4 °C for short-term until the downstream analysis is finished (less than 1 week). After that the samples will be stored at –20 °C

7 Method

1. Prepare the column/tube assembly by inserting a spin column called PrepFiler LySep into a hingeless PrepFiler sample tube. Note that this assembly will work with all the samples except for the bone and tooth where the Bone and Tooth Lysate Tube will be used.
2. Prepare the sample according to Table 5 below.
3. Prepare the PrepFiler lysis solution. Each sample required:
   - Recommended amount of PrepFiler lysis buffer (provided with the kit) shown in Table 6.

Table 5
Samples preparation to be used with PrepFiler® Kits along with the AutoMate Express™ (information for additional sample types is provided in the manufacturer's manuals). Adapted from manufacturer's manual—differences to the manufacturer's recommendations are noted

Sample type	Sample preparation
Blood on FTA or fabric	Add up to 25 mm^2 cutting or punches to the column/tube assembly
Body fluids (semen, saliva) on fabric	Add up to 25 mm^2 cutting or punches to the column/ tube assembly
Body fluids on swab	Add up to 1 swab to the column/tube assembly
Hair	Add up to 5 mm from the root end to the column/tube assembly
Bone and teeth	Add 50 mg of powdered bone or tooth into PrepFiler Bone and Tooth Lysate Tube.
Chewing gum	Chewing gum is prepared by cutting third to half of the gum and then place it in 1.5 ml tube. By using a metal rod the gum is smeared around the inside of the tube. The recommended amount of PrepFiler lysis solution is added and the sample is incubated under the recommended temperature and lysis time (Table 6). The lysate is then transferred to the PrepFiler sample tube (*see* **Note 8**)
Cigarette butts	Remove the first 5 mm paper from the end of the filter. Cut it into small pieces then transfer the pieces to the column/tube assembly

- Recommended amount of 1.0 M DTT (not provided with the kit) shown in Table 6.
- Recommended amount of proteinase K (provided with the kit) shown in Table 6.

4. Pre-digest the samples by adding the recommended amount of PrepFiler lysis solution (Table 6).
5. Incubate the sample under the recommended temperature, lysis time, and shaking speed. Selection of those parameters depends on the sample type (Table 6).

   Note that when the temperature is above 56 °C, no proteinase K is added to the lysis solution. Most proteins denature when the temperature is above 50 °C, so there no need to add proteinase K when the incubation temperature is high (70 °C). A precautionary step has been taken by adding 1.0 M DTT to the solution to maximize the yield of DNA by breaking down any disulfide bonds that may be present in the substrate.
6. Centrifuge the column/tube assembly or the Bone and Tooth Lysate Tube for 2 min and 90 s respectively at 10,000 × *g*.

   In case of the column/tube assembly, if the volume after the centrifugation is less than 300 μl, centrifuge the column/tube assembly for additional 5 min to collect the rest of the

Table 6
Recommended parameters to be used with the AutoMate Express™ PrepFiler® Kits based on the manufacturer's recommendations—(adapted from manufacturer's manual—a variation from the recommendations is noted)

Sample type	PrepFiler lysis solution (µl)	PrepFiler lysis buffer (µl)	1.0 M DTT (µl)	Proteinase K (20 mg/ml)	Recommended incubation temperature and lysis time	Recommended shaking speed
Blood on FTA or fabric	500	500	5	–	At 70 °C for 40 min	750 rpm
Body fluids (semen, saliva) on fabric	500	500	5	–	At 70 °C for 40 min (*see* **Note 9**)	750 rpm
Body fluids on swab	500	500	5	–	At 70 °C for 40 min	750 rpm
Hair	500	500	5	–	At 70 °C for 40 min	750 rpm
Bone and teeth	230	220	3	7 µl	At 56 °C from 2 to 18 h	1100 rpm
Chewing gum	230	220	3	7 µl	At 56 °C for 40 min	750 rpm
Cigarette butts	230	220	3	7 µl	At 56 °C for 40 min	750 rpm

sample and then discard the column. If still less than 300 µl, add PrepFiler lysis buffer to bring the volume to 300 µl.
In case of Bone and Tooth Lysate Tube, transfer the clear lysate to new PrepFiler sample tube. If the lysate solution is less than 150 µl, bring the solution up to 150 µl by adding PrepFiler lysis buffer.

7. Load AutoMate Express sample tubes, elution tubes, tips and holders, and cartridge into the instrument according to manufacturer's instruction.
8. Start the instrument according to manufacturer's instruction.
9. In the instrument:
   First, the sample will be added to the lysis buffer in the cartridge. This lysis buffer contains chaotropic agents that lysis any remain cells, denature the proteins, inhibit nuclease enzymes as well as promote the binding of the DNA to the polymer embedded magnetic beads. Secondly, the magnetic beads will be added to the lysis buffer and the DNA will be bound to the bead and washed to remove impurities. Finally, the DNA is eluted under an alkaline condition using TE buffer. AutoMate Express takes approximately 30 min to finish the extraction. A schematic example of DNA extraction from epithelial cells using AutoMate Express instrument is shown in Fig. 3.

8 Notes

1. In this step, a Phase-Lock Gel tube (Eppendorf, North America) can be used to make the separation cleaner and easier. This technique utilizes gel inside the tube that migrates to between the organic and aqueous phases.
2. We use the spinaroo process to filter the pre-extracted samples from any absorbent material. Spinaroo is a filtration process done by making two holes in the top and bottom of the tube containing the pre-extracted sample, respectively. Then this tube is placed in another tube and centrifuged at 4000 × *g* for 3–5 min. In our laboratory we use a sterile metal needle to make the hole—this is heated in a Bunsen and then easily pierces the plastic tubes.
3. The n-butanol wash is not essential, but it helps to remove any trace phenol that can otherwise act as a PCR inhibitor.
4. DNA in the aqueous phase can be recovered using a standard ethanol precipitation protocol as an alternative; however, we find that that DNA extracts concentrated and washed using Microcon® 100 columns give higher yields and contain less PCR inhibitors.
5. The manufacturer recommends extracting 200 μl of liquid blood directly without predigestion step: we find this produces inhibitors which leads to dropout of some STR markers.
6. If the cellular material is on an absorbent substrate then the lysis buffer should be diluted 1:1 with DNA-free water.
7. Based on our validation study the recommended EZ1 protocol is Trace at all times. When processing absorbent materials using spinaroo followed by the Trace protocol gave higher DNA yield compared with the Tip Dance protocol.
8. Our protocol differs from the manufacturer's recommendation when handling chewing gum. The manufacturer's recommendation: (A) When using Automate Express is to place the gum in a clean petri dish, flatten the piece of gum into a pancake shape of approximately 5 mm thickness. Cover the petri dish and then place it in −80 °C freezer for at least 2 h. Cut and transfer around 50 mg to the column/tube assembly. (B) When using EZ1 instrument is to place 40 mg of chewing gum cut into small pieces to 2 ml EZ1 sample tube.
9. Based on our validation studies the recommended incubation temperature and lysis time for semen samples is 90 °C for 40 min.

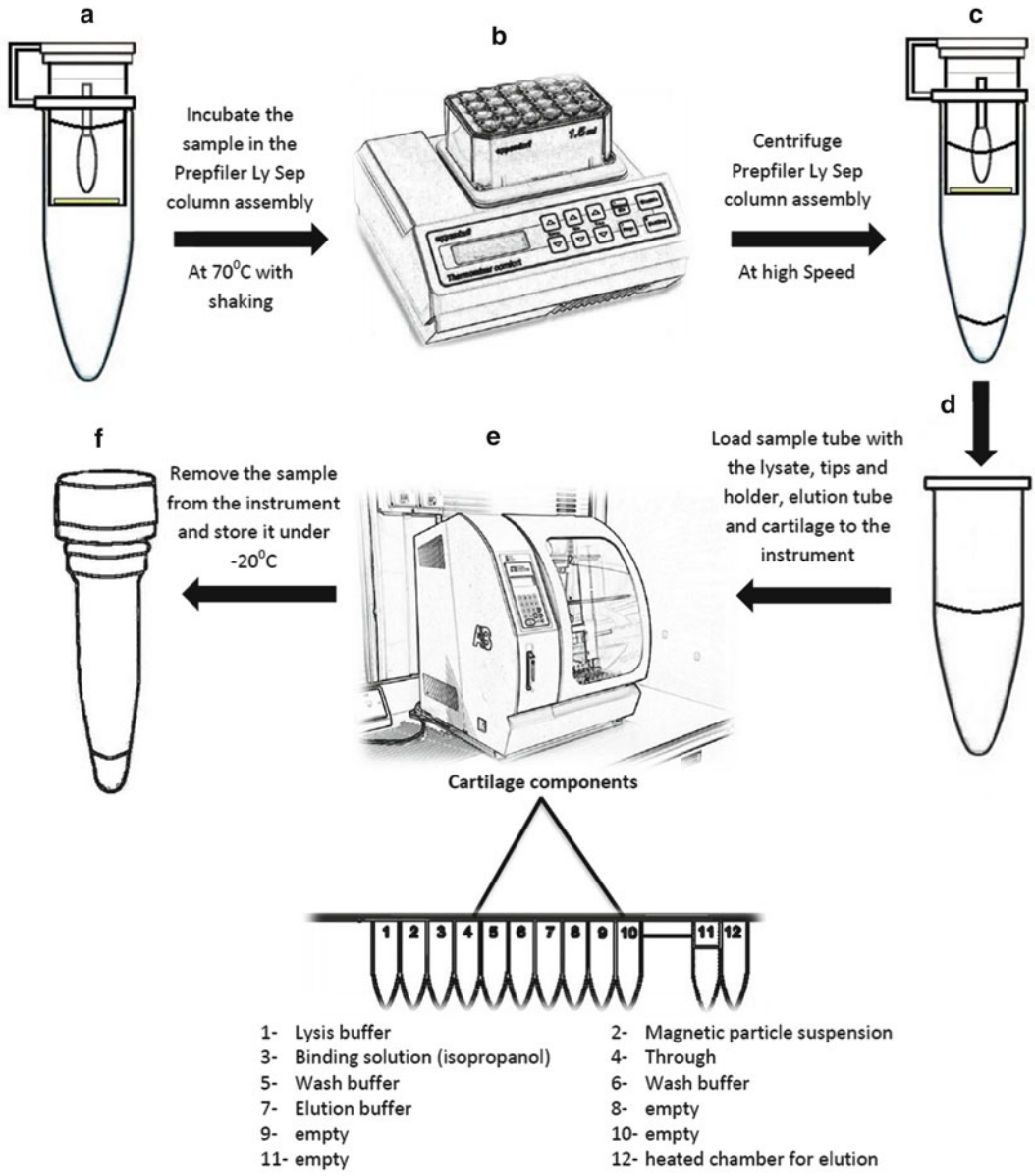


Fig. 3 DNA extraction from epithelial cells (swab) using solid-phase extraction method based on polymer embedded magnetic beads. (**a**) Cellular material is added to PrepFiler LySep column/tube assembly and lysis solution containing PrepFiler lysis buffer and DTT is added to the assembly. (**b**) The sample tube is incubated at 70 °C for 40 min. (**c**) Digested swab is centrifuged for 2 min at 10,000 × *g*. (**d**) Filtrate solution is checked to see if the volume in the sample tube is more than 300 μl if not, the volume will be adjusted to 300 μl with the PrepFiler lysis buffer. (**e**) The sample tube, elution tube, tips and holders, and cartridge are loaded to the AutoMate Express instrument. The extraction process started using the protocol installed in the instrument. (**f**) After the instrument has finished the running protocol each elution tube is closed and stored at fridge for short term until the downstream analysis is finished. After that the samples are stored at −20 °C

References

1. Walsh PS, Metzger DA, Higuchi R (1991) Chelex® 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10:506–513
2. Phillips K, McCallum N, Welch L (2012) A comparison of methods for forensic DNA extraction: Chelex-100 ® and the QIAGEN DNA investigator kit (manual and automated). Forensic Sci Int Genet 6:282–285
3. Stangegaard M, Hjort BB, Hansen TN, Hoflund A, Mogensen HS, Hansen AJ et al (2013) Automated extraction of DNA from biological stains on fabric from crime cases. A comparison of a manual and three automated methods. Forensic Sci Int Genet 7:384–388
4. Amory S, Huel R, Bilic A, Loreille O, Parsons TJ (2012) Automatable full demineralization DNA extraction procedure from degraded skeletal remains. Forensic Sci Int Genet 6:398–406
5. Anslinger K, Bayer B, Rolf B, Keil W, Eisenmenger W (2005) Application of the BioRobot EZ1 in a forensic laboratory. Leg Med 7:164–168
6. Davis CP, King JL, Budowle B, Eisenberg AJ, Turnbough MA (2012) Extraction platform evaluations: a comparison of automate express™, EZ1® advanced XL, and maxwell® 16 bench-top DNA extraction systems. Leg Med 14:36–39
7. Keijzer H, Endenburg SC, Smits MG, Koopmann M (2010) Automated genomic DNA extraction from saliva using the QIAxtractor. Clin Chem Lab Med 48:641–643
8. Montpetit SA, Fitch IT, O'Donnell PT (2005) A simple automated instrument for DNA extraction in forensic casework. J Forensic Sci 50:555–563
9. Jakubowska J, Maciejewska A, Pawlowski R (2012) Comparison of three methods of DNA extraction from human bones with different degrees of degradation. Int J Legal Med 126:173–178
10. Verdon TJ, Mitchell RJ, van Oorschot RAH (2011) Evaluating the efficiency of DNA extraction methods from different substrates. Forensic Sci Int Genet 3:e93–e94
11. Butler JM, Shen Y, McCord BR (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. J Forensic Sci 48:1054–1064

Chapter 6

Extraction of DNA from Skeletal Remains

Suni M. Edson and Timothy P. McMahon

Abstract

Acquisition of DNA from skeletal remains can be a delicate process. With the advent of improved extraction buffers that provide complete demineralization of the osseous materials, extraction of total genomic DNA from nearly any skeletal element is possible. This chapter describes both traditional organic and more newly developed inorganic extraction methods for fresh and dried skeletal remains.

Key words DNA from skeletal remains, Organic extraction, Inorganic extraction, PCIA

1 Introduction

Considering the potentially reactive chemical groups that comprise the molecule, double-stranded DNA is a stable, inert chemical. Buried within the helix, reactive groups are steadied by hydrogen bonds. The bases that make up the DNA itself are protected from the outside by a casing of phosphates and sugars that is reinforced through strong internal stacking forces. DNA's robust packaging protects it from most intracellular decomposition processes, which makes it ideal for use in criminal forensics and human identification.

Although chemically stable, the DNA molecule itself is physically unstable and subject to hydrodynamic shearing forces. DNA in an aqueous medium is a condensed supercoiled molecule that is stabilized by stacking interactions between the individual base pairs, and negative charge repulsions between the phosphate molecules in the DNA backbone. The flow of liquid across the DNA molecule due to pipetting, vortexing, or stirring creates flow resistance across the DNA double strands with enough energy to break the DNA. The longer the DNA molecule, the lower the amount of force needed to break the DNA into smaller fragments. On average any DNA molecule greater than 200 bp is readily susceptible to flow force breakage. However, in criminal and human identity DNA forensic testing, DNA shearing has little to no impact since

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_6,

most Short Tandem Repeat and Mitochondrial sequencing methods test for DNA fragments between 100 and 500 bp.

The greatest impact on the success of forensic human identity testing is degradation of the DNA molecule (DNases, bacteria, body decomposition, etc.) and environmental insults (acidic soil, temperature, humidity, etc.). 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.

To extract DNA from cells, four primary extraction techniques are available to DNA forensic laboratories. These include organic, Chelex®, FTA®, and solid phase (inorganic) methods. The extraction method chosen will depend upon the biological sample being examined as well as what environmental or chemical insults the sample may have been exposed to. Laboratories can chose to have a combination of different extraction methods validated for use to insure the greatest chance of success. The scope of this chapter will deal with organic and solid phase extractions.

Organic extraction, commonly referred to as Phenol-Chloroform or PCIA extraction, has been used for DNA purifications since the late 1950s. Although time-consuming, with many transfer steps and the requirement to use harmful chemicals, organic extractions are still the gold standard to which all new extraction methods are compared. Organic extractions start with the addition of a lysis buffer that contains a buffer agent (commonly Tris), a detergent (SDS or *N*-Lauroylsarcosine), Proteinase K, and a chelating agent (EDTA or EGTA). The detergent and Proteinase K are used to solubilize the cellular membrane and denature the proteins that protect the DNA in the nucleus. Proteinase K, which is necessary for efficient protein denaturation, is optimally active in 0.5–1.0 % detergent and 56 °C.

Once released, the DNA is susceptible to DNase activity. The addition of EDTA inhibits DNase activities by binding divalent cations like Mg^{++}. After digestion, Phenol:Chloroform:Isoamyl Alcohol (24:24:1) is added at 50:50 ratio to extraction buffer. Phenol is an organic solvent that has a specific gravity of 1.07 and forms the lower organic phase when mixed with an aqueous solution and acts as a protein solvent. Chloroform is an organic solvent that acts as a protein and RNA solvent, while Isoamyl Alcohol functions as a foam reducing agent. The Lysate/PCIA mixture is vortexed until an emulsion is formed and then separated into an aqueous (top layer) and organic (bottom layer) phase by centrifugation. The denatured proteins and cellular debris are pulled into the organic layer, the lipids will accumulate at the interface between the aqueous and organic layer, and the DNA will accumulate in the

aqueous phase. Due to the chemical properties of PCIA, it is essential to denature the protein away from the DNA and to buffer the solution to a pH > 7.8, in order to prevent DNA from accumulating in the organic layer. After several PCIA washes, the aqueous phase can be extracted with Chloroform or *n*-Butanol to remove any residual traces of Phenol, a potent inhibitor of downstream amplification process. Then the DNA can be purified and concentrated by either ethanol precipitation or centrifugal filter units/ultrafiltration concentrators.

However, recent advances in solid phase (inorganic) methods and ultrafiltration concentrators have allowed DNA to be purified from lysate without the need for the PCIA purification steps, which decreases processing times, limits the number of transfer steps, and removes interactions with hazardous chemicals.

It is the authors' desire to outline procedures for the organic and inorganic extractions of dried and aged skeletal material and a modified inorganic procedure for fresh skeletal material. The procedures outlined below use Extraction (demineralization) buffer (0.5 M EDTA, pH 8.0; 1 % *N*-Lauroylsarcosine). The high amount of EDTA serves two functions, one to inhibit DNase activity, and two, to completely dissolve the CA^{++} rich bone matrix and free any and all DNA that may be contained in challenged bone samples [1]. However, the demineralization extraction buffer and procedures outlined below can be used on any biological specimens to obtain higher yields of DNA, when compared quick lysis and purification methods such as Chelex or FTA.

2 Materials

All materials may be stored at room temperature unless otherwise noted.

Recommendation: Those objects that can be ultraviolet (UV) irradiated should be prior to initiating the protocol. Irradiation time, which will vary depending on the equipment used, should be set to deliver 6.0 J/cm^2. These include such items as the Waring blender cups, 50 ml and 15 ml conical tubes, and 1.7 ml microcentrifuge tubes. Pipettes should never be UV irradiated, as repeated exposure to UV light will cause the plastic to decay; however, sterilization by wiping down the outside with 8.5 % bleach (v/v) (70 mM NaOCl) or other DNA removal solution is recommended. Certain extraction reagents whether purchased externally or made internally may be UV irradiated, but the contents of the QIAquick and MinElute kits should never be UV irradiated.

To reduce the chance of contamination, the surface of all hoods (laminar flow, chemical fume, and PCR) and bleach tolerant equipment should be wiped down with 8.5 % (v/v) commercial bleach followed by a 95 % ethanol wipe to reduce the corrosive impact of

the bleach. Additionally, staff participating in the extraction, to include all individuals entering the lab, should, at a minimum, wear the following personal protective equipment (PPE): non-permeable, disposable laboratory coat; one layer of latex or nitrile gloves; non-permeable, disposable sleeves; goggles; and a face mask. Gloves and sleeves should be changed between samples during cleaning in order to reduce contamination. During the preparation of the dried skeletal sample, the wearing of two pairs of gloves is recommended, as it is fairly common to damage the outer pair of gloves during cleaning. Individuals with longer hair should pull hair back away from the face. Hair nets/caps are not required, but can be used. Coats should be discarded at the end of the day, or sooner, if the scientist feels that the coat has been compromised.

2.1 Skeletal Sample Preparation and Cleaning (Dried)

1. Dremel® rotary tool (Bosch, Stuttgart, Germany).
2. Fit for use aluminum oxide sanding bits (size dependent on need) and emery grinding wheels compatible with the Dremel® rotary tool.
3. Surgical/dental mallet and osteotome/periodontal chisel.
4. Parafilm® M Barrier Film.
5. 50 ml polypropylene conical tubes (BD, Franklin, NJ, USA).
6. Absolute ethanol—99.8 % (Pharmco-AAPER, Brookfield, CT, USA).
7. Waring blender 700S/700G with appropriately sized mini container (MC1, MC2, or MC3) (Waring, Torrington, CT, USA).
8. 15 ml polypropylene conical tubes (Sarstedt, Newton, NC, USA).
9. Liquinox (Alconox, White Plains, NY, USA).

2.2 Intact Tooth Preparation and Cleaning (Dried)

1. Absolute ethanol.
2. 50 ml polypropylene conical tubes.
3. Small sonicating water bath.
4. Dental Handpiece (Forza L50K Lab Micromotor, Brasseler USA, Savannah, GA, USA).
5. #2, #4, and #6 dental burs.
6. Periodontal chisel.
7. Forceps or spoon excavator.
8. 15 ml polypropylene conical tubes.
9. 4 × 4 mm sterile gauze pad.

2.3 Organic Extraction of Powdered Skeletal Samples or Teeth (Dried)

1. 15 ml polypropylene conical tubes (*see* **Note 1**).
2. Extraction (demineralization) buffer: 0.5 M EDTA, pH 8.0; 1 % *N*-Lauroylsarcosine.
3. Proteinase K (20 mg/ml).

4. Incubator shaker capable of maintaining 56 °C.
5. Phenol:Chloroform:Isoamyl Alcohol (25:24:1), pH8.0 ± 0.2.
6. Centrifuge.
7. *N*-Butanol.
8. Amicon® Ultra-4/30 K centrifugal filter units (Millipore, Billerica, MA, USA).
9. TE Buffer (10 mM Tris, 1 mM EDTA, pH 7.5) (aka TE^{-4}).
10. Costar® 1.7 ml polypropylene microcentrifuge tubes (Corning, Ithaca, NY, USA).

2.4 Nonorganic Extraction for Both Skeletal Samples and Teeth (Dried)

1. 15 ml polypropylene conical tubes.
2. Extraction (demineralization) buffer: 0.5 M EDTA, pH 8.0; 1 % *N*-Lauroylsarcosine.
3. Proteinase K (20 mg/ml).
4. Parafilm® M Barrier Film.
5. Incubator shaker capable of maintaining 56 °C.
6. Amicon® Ultra-4/30 K centrifugal filter units.
7. Costar® 1.7 ml polypropylene microcentrifuge tubes.
8. QIAquick PCR Purification Kit (QIAGEN, Hilden, Germany).
   (a) PB Buffer
   (b) PE Buffer
   (c) EB Buffer (provided, but not used)
   (d) QIAquick spin columns
9. Microcentrifuge.
10. 95–100 % ethanol.
11. TLE (10 mM Tris, 0.01 mM EDTA; pH 7.5).

2.5 Skeletal Sample Preparation and Cleaning (Fresh)

1. 15 ml polypropylene conical tubes.
2. Scalpels.
3. 95 % ethanol.
4. Mortar.
5. Surgical/dental mallet and osteotome/periodontal chisel.
6. Waring blender 700S/700G with appropriately sized mini container (MC1, MC2, or MC3).

2.6 Intact Fresh Tooth Preparation and Cleaning (Fresh)

1. 15 ml polypropylene conical tubes.
2. Scalpels.
3. 95 % ethanol.
4. Mixer/Mill MM 200 (Retsch, Haan, Germany).

2.7 Nonorganic Extraction for Both Skeletal Samples and Teeth (Fresh)

1. 15 ml polypropylene conical tubes.
2. Extraction (demineralization) buffer: 0.5 M EDTA, pH 8.0; 1 % *N*-Lauroylsarcosine.
3. Proteinase K (20 mg/ml).
4. Incubator shaker capable of maintaining 56 °C.
5. Amicon® Ultra-4/50 K centrifugal filter units (Millipore, Billerica, MA, USA).
6. MinElute PCR Purification Kit (QIAGEN, Hilden, Germany).
   (a) PB Buffer
   (b) PE Buffer
   (c) EB buffer
   (d) MinElute columns
7. Costar® 1.7 ml polypropylene microcentrifuge tubes.
8. Microcentrifuge.
9. 95–100 % Ethanol.

3 Methods

All steps may be carried out at room temperature unless otherwise noted.

3.1 Sample Selection

Prior to extraction, selection of the optimal bone sample for testing will increase the odds of obtaining reportable results. If possible, collaboration with an anthropologist is recommended prior to cutting a sample for extraction from the intact bone. In general, compact bones with a dense physical structure, such as the femur, tibia, and humerus, tend to provide greater yield of DNA [2–4]. However, modified extraction protocols can negate the impact of the sample type itself, and almost any skeletal element should provide sufficient quality DNA [1, 5].

Dental elements should be selected with care. Compromised external structure may introduce bacteria or other factors of decay following death of an individual. Pre-mortem dental work or disease may likewise compromise the internal integrity of the tooth structure and eliminate or reduce the recovery of DNA [6, 7].

3.2 Skeletal Sample Preparation and Cleaning (Dried)

This description is for any dried osseous element other than teeth. For tooth preparation, see below.

The following steps should take place in a bone sanding hood or other laminar flow hood with sufficient ventilation (*see* **Note 2**).

1. Sand the exterior surface of the bone (*see* **Note 3**) using a clean aluminum oxide sanding bit fitted into a Dremel® rotary tool. All visible surfaces of the bone need to be sanded so as to

remove any dirt, vegetative materials, or other exogenous contaminants. All trabecular (spongy) bone should be sanded away. Spongy bone can harbor detritus that may either contaminate or inhibit the extraction. In instances where visible trabecular bone is present between layers of compact bone (such as the cranial vault), it is useful to use a cutting wheel to separate the layers of compact bone in order to remove the spongy bone prior to proceeding. Retention of trabecular bone that may have been exposed to the environment for an extended period of time will increase the chance of introducing unwanted materials to the extraction.

2. Remove approximately 0.2–0.5 g of bone specimen from the larger, now sanded sample, using either a sanding wheel or a mallet and chisel of appropriate size. The remaining portion of bone sample can be repackaged by wrapping in Parafilm or placed in a sterile conical tube. The remaining sample can thus be retained for further processing, returned to the submitting agency, or place in long-term storage. The following steps should take place in a laminar flow hood (*see* **Note 2**).
3. Place the bone fragment to be extracted in a 50 ml conical tube containing approximately 25 ml sterile deionized water (diH_2O). By hand, shake the tube vigorously back and forth several times.
4. Decant the water into a waste container.
5. Repeat **steps 4** and **5** until the water is no longer cloudy (*see* **Note 4**).
6. Cover the bone fragment in the conical tube with absolute EtOH. Shake the container back and forth vigorously several times and decant liquid into a waste container.
7. Decant bone sample into a cleaned weigh boat or other nonporous, wide-mouthed container, and allow the sample to completely dry prior to proceeding. Depending on the density of the bone, this may take from 1 to 2 h.
8. Once the sample is dry, place within a small sterilized blender cup (*see* **Note 5**). Seal lid prior to removing from the hood and placing on the blender base.
9. "Blend" bone sample until a fine powder is generated (*see* **Note 6**).
10. Return the blender cup to the hood prior to removing the lid. Pour powdered bone into a cleaned weigh boat, and then transfer ~0.2 g to a sterile 15 ml conical tube. Any remaining powder should be transferred to a separate 15 ml conical tube for storage (*see* **Note 7**).

The procedure can be paused at this point and the bone powder stored at –20 °C.

3.3 Intact Tooth Preparation and Cleaning (Dried)

This description is for intact dried teeth. Powdered teeth may be extracted using either of the following methods with no other preparation involved. Ideally, a laboratory would collect enough tooth powder to perform duplicate extractions.

1. Examine the exterior of the tooth. If the tooth is whole and undamaged, proceed to **step 3**.
2. If the tooth is cracked, damaged, or has untreated caries, the exterior of the tooth should be cleaned using a 4×4 sterile gauze pad moistened with 8.5 % (v/v) commercial bleach. When the exterior surface is clean, immediately remove any remaining bleach using a 4×4 sterile gauze pad moistened with absolute EtOH. Proceed to **step 5**.
3. Place the whole, undamaged tooth in a 50 ml sterile conical tube and cover with 25 ml of 8.5 % (v/v) commercial bleach. Place sealed tube in a sonicating water bath for 5 min.
4. Remove tooth from bleach wash and wipe down with first a 4×4 sterile gauze pad moistened with 8.5 % (v/v) commercial bleach and then second a sterile gauze pad moistened with absolute EtOH.
5. Place cleaned tooth in a cleaned weigh boat and allow to dry under UV irradiation in a laminar flow hood for approximately 15 min. If the tooth is not dry at the end of 15 min, turn off the UV light and allow to dry completely before proceeding.

   The following steps should take place in a bone sanding hood or other laminar flow hood with sufficient ventilation (*see* **Note 2**).
6. Using a #2 or #4 dental bur, slowly cut around the base of the crown of the tooth. The intact tooth can be held with a hemostat if necessary; however, holding the tooth by hand allows for easier manipulation during cutting. Avoid cutting the enamel or cutting so low on the roots that the roots cannot be removed in a single piece. Cut around the crown until approximately 1 mm is left intact. This cutting needs to be performed over a clean weigh boat. The powder recovered from the external cut should be saved in a sterile 15 ml conical tube, though not used for extractions unless necessary.
7. Place a small, periodontal chisel between the crown and the roots and twist very gently to remove the crown from the root (*see* **Note 8**).
8. Should there be visible pulp, remove it to a clean 15 ml conical tube using a spoon excavator or forceps.
9. Using a #4 or #6 dental bur, carefully drill the dentin from the interior of the crown and root, catching it in sterile weigh boat (*see* **Note 9**). Remove as much dentin as possible, being careful to not puncture either the root or crown.

10. As drilling progresses, regularly remove the powder generated to a UV irradiated 15 ml conical tube. When approximately 0.2 g of powder has been collected, switch to a second UV irradiated 15 ml conical tube and collect another 0.2 g of powder, or until no dentin remains. If the tooth is too small, as may be the case for incisors, only 0.2 g total dentin powder may be recovered.
11. Store the remaining tooth structure in a UV irradiated 15 ml conical tube to prevent further damage.

The procedure can be paused at this point and the bone powder stored at –20 °C.

3.4 Organic Extraction for both Dried Skeletal Samples and Teeth (Dried)

The following steps should take place in a laminar flow hood.

1. Start with approximately 0.2 g of powdered osseous material or dentin. A reagent blank should be initiated at this step and carried through the remainder of the procedure.
   (a) If starting with 0.3–0.5 g of either, the volume of some reagents will need to be adjusted accordingly.
   (b) If starting with less than 0.2 g, the procedure may proceed as indicated with a potential reduction in the final volume recovered.
2. Add 3 ml of extraction buffer and 100 µl of proteinase K (proK) to each tube. The buffer may be prewarmed to 56 °C if desired.
   (a) If starting with 0.3–0.5 g of powder, maintain the same volume of extraction buffer, but increase the proK to 200 µl.
3. Shake or invert the tubes gently to completely saturate the bone powder. Continue to gently shake until no dry spots are visible in the powder.
4. Place the tubes into an incubator/shaker set to 56 °C. The tubes should be set to an angle of approximately 45° and gently agitated overnight. Ensure that the liquid does not touch the cap of the tube (*see* **Note 10**).

   The following steps should take place in a chemical fume hood.
5. Add 3 ml of phenol:chloroform:isoamyl alcohol (PCIA) to each tube.
6. Mix vigorously until a complete emulsion is formed.
7. Centrifuge tubes for 3 min at 4000 × *g*. There should be a clear delineation between the layers.
8. Transfer the aqueous (top) layer of each sample to clean 15 ml conical tubes (*see* **Note 11**).
9. Repeat **steps 5–8** until the interface is clean (or a minimum of two times) (*see* **Note 12**).

10. Add 3 ml of *n*-Butanol to each tube.
11. Mix thoroughly.
12. Centrifuge tubes for 3 min at 4000 × *g*. Again, there should be a clear delineation between the layers.
13. Remove most of the upper layer to a waste container. This will aid in cleanly removing the desired aqueous (bottom) layer.
14. Remove the bottom layer of each sample to clean Ultra-4/30 K centrifugal filters. Take care not to transfer any remaining butanol along with the sample/reagent blank. Butanol may cause holes to form in the filter membranes, and encourage loss of DNA.
15. Spin the filters for 40–50 min at 2000 × *g* (*see* **Note 13**). There should be approximately 200 μl of sample remaining at this time. If there is markedly more volume left, the filters should be spun for additional time until this volume is reached (*see* **Note 14**).
16. Discard the filtrate.
    The following steps should take place in a hood. No specific requirements for type of hood.
17. Add 2 ml of sterile TE^{-4} Buffer to each filter unit.
18. Spin all filter units for 10–15 min at 2000 × *g*. The volume of the retentate should again be approximately 200 μl. If it is not, return the filter units to the centrifuge for an additional time until such volume is reached.
19. Discard the filtrate.
20. Repeat **steps 17–19** once.
21. Recover the retentate and transfer to clean 1.7 ml tubes. The tip of the P-100 pipette may not fit all the way to the bottom of the filter unit. If so, recover the majority of the sample with a P-100 and the remaining sample with a P-10.
22. Measure the final volume and bring to 200 μl with TE buffer as needed. If the bone/tooth sample started at a significantly lower powder weight (less than 0.1 g) and is deemed to be of poor quality, it is recommended that the final volume be brought only to 100 μl.
23. Sample is now ready for processing.

3.5 Nonorganic Extraction for Both Skeletal Samples and Teeth (Dried)

The following steps should take place in a laminar flow hood.

1. Start with approximately 0.2 g of powdered osseous material or dentin. If starting with 0.3–0.5 g of either, the volume of some reagents will need to be adjusted accordingly. If starting with less than 0.2 g, the procedure may proceed as indicated with a potential reduction in the final volume recovered. A reagent blank should be initiated at this step and carried through the remainder of the procedure.

2. If extracting 0.26 g of bone/tooth powder or less, add 4 ml of extraction buffer to the sample and the reagent blank. If extracting more than 0.26 g of bone/tooth powder, add 7.5 ml of extraction buffer to the sample and the reagent blank. The buffer may be prewarmed to 56 °C if desired.
3. Add 200 µl proteinase K to each tube.
4. Shake or invert the tubes gently to completely saturate the bone powder. Continue to gently shake until no dry spots are visible in the powder.
5. Place the tubes into an incubator/shaker set to 56 °C. The tubes should be set to an angle of approximately 45° and gently agitated overnight. Ensure that the liquid does not touch the cap of the tube (*see* **Note 15**).
6. Centrifuge tubes for 3 min at 4000 × *g*. This will bring any remaining bone powder to the bottom of the tube.

   The following steps should take place in a laminar flow hood.
7. Transfer up to 4 ml of supernatant to Ultra-4/30 K centrifugal filter unit.
8. Spin the filter unit for 40–60 min at 2000 × *g* (*see* **Note 13**).
9. Discard filtrate.
10. If starting with more than 4 ml of supernatant, add the remaining volume to the appropriate filter unit and spin for an additional 40–60 min at 2000 × *g*.
11. Repeat **steps 7–10** until all of the supernatant has been added to the filter unit.
12. The final retentate volume should be approximately 120 µl (*see* **Note 16**). If this is not so, the filter unit may be spun for additional time at 2000 × *g*, until a final volume of 250 µl or less is reached.
13. Remove the retentate from the filter unit directly to a 1.7 ml microcentrifuge tube. The tip of the P-100 pipette may not fit all the way to the bottom of the filter unit. If so, recover the majority of the sample with a P-100 and the final volume with a P-10 (*see* **Note 17**).
14. Measure and record the recovered volume.
15. Add 5 volumes of PB Buffer to 1 volume of sample. For example, if the recovered volume of the sample was 100 µl, 500 µL of PB Buffer would be added.
16. Mix well and tap down to remove any liquid from the lid of the tube.
17. Assemble the appropriate number of QIAquick spin columns in the provided 2 ml collection tubes.
18. Aliquot up to 750 µl of the buffer/sample mixture into each QIAquick spin column.

19. Spin columns for 30 s at 17,900 × *g* in a microcentrifuge. If there is still visible liquid on the membrane after this step, spin columns for an additional 30 s at 17,900 × *g*.
20. Discard waste (*see* **Note 18**).
21. Repeat **steps 18–20** until all of the sample has been added to the spin column.
22. Add 750 μl PE buffer to each spin column (*see* **Note 19**).
23. Spin columns for 30 s at 17,900 × *g* in a microcentrifuge. If there is still visible liquid on the membrane after this step, spin columns for an additional 30 s at 17,900 × *g*.
24. Discard waste.
25. Centrifuge the spin columns for an additional 60 s at 17,900 × *g*.
26. Place spin column in new, clean 1.7 ml microcentrifuge tubes.
27. Add 100 μl of sterile TLE to the center of the column
28. Let stand for at least 1 min.
29. Centrifuge the column for 1 min at 17,900 × *g* (*see* **Note 20**).
30. If needed, transfer the eluate to a new 1.7 ml microcentrifuge tube and discard the spin column.
31. Add 500 μl of PB Buffer to the 100 μl of eluate.
32. Repeat **steps 16–26**.
33. Add 50–200 μl of TLE to the center of the column and allow to stand for at least 1 min (*see* **Note 21**).
34. Spin the columns for 1 min at 17,900 × *g*.
35. Transfer eluate to a new, clean microcentrifuge tube as needed and discard column.
36. Sample is now ready for processing.

    Extracts may be held at –20 or –80 °C for extended storage. In the short term, 4 °C is adequate. It is best to minimize freeze/thaw cycles as freezing may damage the DNA.

3.6 Skeletal Sample Preparation and Cleaning (Fresh)

Unless the bone specimen is suspected of being exposed to chemical or other agents, there is no need to perform the following steps in a hood. However, a Biological Safety Cabinet (BSC) hood may be used at the discretion of the scientist.

1. Remove any tissue or debris that might be adhering to the sample using a scalpel. Depending on the source or the needs of the laboratory, the removed tissue may be stored in a 15 ml conical tube for evidence or extraction.
2. Place the sample in a 50 ml conical tube and add enough 95 % EtOH to cover the sample.
3. Shake the tube vigorously.
4. Decant off the ethanol.

5. Repeat **steps 2–4** twice.
6. Place sample in a dry, clean weigh boat and allow to dry in a laminar flow hood. Sample must be completely dry before proceeding. This should take approximately 1 h.
7. Place sample in a clean mortar and cover the mortar with Parafilm (*see* **Note 22**).
8. Using a surgical mallet and an osteotome, punch a small hole in the Parafilm, and split the bone sample into fragments. Collect approximately 1.0 g of bone fragments. The remainder of the bone, if there is any, can be placed in a 50 ml conical tube for storage.
9. Place 1.0 g of bone fragments into a Waring blender cup and seal the lid carefully (*see* **Note 5**).
10. Place the blender cup on a blender base and "blend" the bone sample until a fine powder is generated (*see* **Note 6**).
11. Pour the powder into a clean weigh boat and transfer to a 15 ml conical tube (*see* **Note 23**).

    The procedure can be paused at this point and the bone powder stored at –20 °C.

3.7 Intact Fresh Tooth Preparation and Cleaning (Fresh)

Unless the tooth specimen is suspected of being exposed to chemical or other agents, there is no need to perform the following steps in a hood. However, a Biological Safety Cabinet (BSC) hood may be used at the discretion of the scientist.

1. Place an intact tooth in a clean weigh boat.
2. Cover the tooth with 95 % EtOH.
3. Using a scalpel, remove any adhering dirt or tissue.
4. Wipe down the exterior of the tooth thoroughly with a 4 × 4 sterile gauze pad moistened with 8.5 % (v/v) commercial bleach, followed by sterile gauze pad moistened with EtOH.
5. Place clean tooth in a cleaned weigh boat and allow to dry for at least 30 min or until completely dry.
6. Place the entire tooth in the Mixer/Mill jar along with the ball. Seal the lid in place and follow manufacturer's instructions for setting the jar in place in the Mixer/Mill itself (*see* **Note 24**).
7. Turn the Mixer/Mill to the following settings: Frequency – $1/S = 25.0$; Time = 30–45 s.
8. Pulverize the tooth.
9. Pour the tooth powder into a clean weigh boat and the transfer to a clean 15 ml conical tube.
10. If significantly more than 1.0 g of powder is recovered, the remaining powder should be transferred to a second 15 ml conical tube for storage.

The procedure can be paused at this point and the powder stored at –20 °C.

3.8 Nonorganic Extraction for Both Skeletal Samples and Teeth (Fresh)

1. Start with approximately 1.0 g of bone or tooth powder.
2. Add 3.0 ml prewarmed extraction buffer and 100 µl proteinase K (*see* **Note 25**).
3. Mix thoroughly by shaking the tubes until there are no dry patches of powder in the tube, particularly the very bottom.
4. Place the tubes into an incubator/shaker set to 56 °C. The tubes should be set to an angle of approximately 45° and gently agitated overnight (*see* **Note 26**). Ensure that the liquid does not touch the cap of the tube.
5. Centrifuge tubes for 3 min at 4000×*g*. This will bring any remaining bone powder to the bottom of the tube.

   The following steps should take place in a laminar flow hood.
6. Transfer the supernatant to clean Ultra-4/50 K centrifugal filters. Take care to not aliquot any bone powder as this will tend to clog the filter unit.
7. Spin the centrifugal filters in a centrifuge at 2700×*g* for approximately 60 min (*see* **Note 13**). The final volume should be 120v µl or less (*see* **Note 14**).
8. Remove the retentate from the filter unit directly to 1.7 ml microcentrifuge tubes. The tip of the P-100 pipet may not fit all the way to the bottom of the filter unit. If so, recover the majority of the sample with a P-100 and the final volume with a P-10.
9. Measure and record the recovered volume (*see* **Note 17**).
10. Add 5 volumes of PB Buffer to 1 volume of sample. For example, if the recovered volume of the sample was 100 µl, 500 µl of PB Buffer should be added.
11. Mix well and tap down to remove any liquid from the lid of the tube.
12. Place clean MinElute columns into the 2.0 ml collection tubes.
13. Transfer the DNA/PB Buffer solution to the columns.
14. Centrifuge columns at 17,900×*g* for 1 min.
15. Discard filtrate and return column to the same tube (*see* **Note 18**).
16. Add 750 µl PE Buffer to each column and incubate at room temperature for 5 min (*see* **Note 19**).
17. Spin columns at 17,900×*g* for 1 min.
18. Discard filtrate and return MinElute column to the same collection tube.
19. Spin columns at 17,900×*g* for 1 min.

20. Place column in a clean 1.7 ml microcentrifuge tube (*see* **Note 20**).
21. Add half of the final eluate volume of EB Buffer to the columns and incubate for 1 min at room temperature. The final target volume is typically 50 μl; however, it can be as little as 10 μl.
22. Centrifuge columns for 1 min at 17,900×*g*.
23. Repeat **steps 21** and **22**.
24. Samples are now ready for quantification and/or amplification.

    Extracts may be held at −20 or −80 °C for extended storage. In the short term, 4 °C is adequate. It is best to minimize freeze/thaw cycles as freezing may damage the DNA.

4 Notes

1. While the brand and size of 15 ml conical tube used for PCIA extraction can be chosen by the laboratory, the tube must be composed of polypropylene. Other materials, such as polystyrene, will dissolve when exposed to phenol.
2. Sanding of the fragment should occur within a bone sanding hood or laminar flow hood. It is critical for the powder generated from the sanding of the bone to be captured within the hood or removed by a ducted vacuum system. The powder generated by sanding is very fine and there is the potential for cross contamination of samples within the laboratory.
3. The element to be sampled is usually chosen by an anthropologist or a medical examiner prior to extraction. It is unusual for the DNA bench scientist to be allowed the choice of element to be processed for DNA. It is optimal for the osseous fragment submitted for DNA processing to be at least 0.5 g. Fragments smaller than these are difficult to hold during the cleaning process.
4. The purpose of the washing step is to remove any dirt and debris that may still be remaining on the sample as well as any bone powder that may be on the bone. Washing, along with the sanding procedure, reduces the possibility of recovering exogenous DNA during the extraction procedure. Exogenous DNA can cause mixtures with the endogenous DNA or even overwhelm the authentic profile [8].
5. Equipment used to pulverize the bone sample should be cleaned thoroughly between uses. It is recommended that the blender cup be cleaned with at least one wash of each of the following liquids in order: 1 % liquinox with water, 8.5 % (v/v) commercial bleach, water, and 95 % EtOH; and then exposed to UV irradiation. Cups should be completely dry before grinding of the samples.

6. Rather than removing the blender cup to the hood to determine the degree of pulverization, place a gloved hand on top of the blender cup while the motor is running. Larger pieces of bone may be felt as they bounce against the rubber lid. Take care not to "over blend" the sample. Excessive heat may damage the endogenous DNA. If you believe the bone powder is becoming hot, you may turn off the blender, let it cool, and turn the blender back on. It is also possible for the bone sample to become lodged under the blades of the blender cup. If this occurs, stop the blender and attempt to remove the sample from under the blades by tapping on the counter or rotating the blades from below. If this is insufficient to dislodge the sample, remove the lid of the blender cup (in a hood if dealing with a dried specimen) and manually dislodge the bone using either a periodontal chisel or forceps.
7. When using a Waring blender, the entirety of the powder will not be extremely fine. Another process or tool, such as a Freezer/Mill (SPEX, Metuchen, NJ, USA) or Mixer/Mill, may be used for powdering. When transferring the powder to use in the 15 ml conical tube, you may decant the larger pieces into a different conical tube for storage and then transfer the finer powder into a tube for extraction. It is also possible to use more than 0.2 g of bone powder for extraction. Up to 0.5 g of powdered bone may be used. More than 0.5 g of bone is not recommended as there will be a marked decrease in the dissolution of the powder in the demineralization buffer and an increase in inhibition.
8. The point in having a small notch in the tooth crown and root is to enable correct reassembly of the tooth. While the cementum of the roots contains perhaps the best source of DNA in the tooth [7, 9], acquiring this tissue requires destruction of the tooth root. The method described herein is designed to minimize external tooth damage so as to be able to return an intact tooth structure to a family member of a missing person. (After removal of the dentin, the tooth structure can be glued back together.) If structural integrity of the tooth is not an issue, skip **steps 9** and **10**. The root can be placed into the finger of a latex or nitrile glove and crushed with a hammer or pulverized using a Mixer/Mill (as described in the fresh tooth preparation) or equivalent. The crown can be stored as in **step 11** and then proceed as normal.
9. There is a great deal of static electricity generated during the drilling of the tooth. It is a challenge to dissipate this energy without losing some of the powder being drilled. To ameliorate this issue, a large beaker of water can be set to steam within the laboratory, preferably within a few feet of the processing hood. Keeping a utility wipe such as a Kimwipe (Kimberly-Clark,

Neenah, WI, USA) moistened with 8.5 % (v/v) bleach solution nearby to regularly moisten the fingertips and prevent charge build-up is also helpful; however, care should be taken to not introduce bleach to the powder.

10. The 0.2 g bone/tooth powder will dissolve in approximately 8 h. If time is of the essence, observe the tubes during the incubation process. Once the powder has been completely dissolved, you may proceed to the organic extraction steps of this protocol. However, it should be noted that it is not an infrequent occurrence that some bone powder will remain in the solution even after an overnight incubation.
11. It is crucial that the PCIA solution be maintained at pH > 7.8. Depending on the source of the PCIA being used, a separate buffer will be supplied to equilibrate the solution pH. In a high volume laboratory, PCIA stored under normal conditions will usually be exhausted before any significant oxidation occurs, which will change the pH of the solution. However, low volume labs may use a single bottle of PCIA for several months and there is the potential for oxidation to occur. PCIA that has been oxidized beyond usefulness will appear yellow or red and in these instances if the pH < 5. If this occurs the DNA will accumulate in the organic phase and not the aqueous phase during purification.
12. When drawing off the aqueous layer after each PCIA wash, take care not to collect any of the lipid-protein interface or organic phase. This is especially true in the final wash, as the proteins and other waste at the interface can inhibit downstream amplification processes. Some DNA is lost each time that a PCIA wash is performed and therefore it is prudent to minimize the number of PCIA washes if possible.
13. When placing the centrifugal filter units/ultrafiltration concentrators in the centrifuge, take care to align the largest parts of the unit with the center column of the centrifuge and the outer wall. This will allow a more efficient flow through of the wash waste. If the filtrate seems to flow through faster or more completely than is to be expected, there is a chance that there is a hole in the filter membrane. In this case, recover the filtrate and transfer it to a new, clean filter unit and continue with the centrifugation steps.
14. In some instances it will take significantly longer than the indicated times for the waste product to flow through the filter. This is due to either an excessive amount of high-quality DNA being present in the solution or extra waste product. It is possible for the filter unit to become completely clogged and no TE will flow through. If this occurs, the retentate should be transferred to a second clean filter unit for the additional washes.

15. With the larger volume of liquid, there may be concern that there will be leakage from the cap of the tube. Bone or tooth powder trapped in the threads of the cap will increase this possibility. Therefore, to prevent loss of extract, Parafilm may be wrapped around the top of the tube after capping. Scissors will most likely be needed to remove the wrap after incubation.
16. The maximum input of extract into a QIAquick column is 250 μl. It is optimal to reach a volume of retentate that is equal to or less than that during **steps 7–12**. However, it is possible to proceed if the retentate volume is greater than that. Additional QIAquick columns will need to be used for any volume over 250 μl and the final product recovered pooled.
17. It should be noted that the retentate will be somewhat viscous. It is rather difficult to pipette without some level of bubbling. Do not over-handle the liquid as the downstream steps require a reasonably accurate measurement of retentate volume. If too many bubbles form, they can be brought down by gently tapping the tube on the counter or popping them with a P-10 tip.
18. It is critical that the flow-through is discarded after spin. Otherwise the liquid will not pass through the column during the following spin.
19. PE Buffer is received in the kit as a concentrate. The buffer will need to be prepared in advance of processing as indicated in the manufacturer's instructions using 96–100 % EtOH.
20. When centrifuging the columns in the 1.7 ml microcentrifuge tubes, it is likely that the caps of the microcentrifuge tubes will be ripped off by the centrifuge. It is best to have several clean 1.7 ml microcentrifuge tubes available for downstream processing and storage. If you are spinning this particular set of tubes more than once, it is not necessary to transfer the eluate to a new microcentrifuge tube until the process is complete.
21. The volume of TLE added at this step will depend on the downstream processing of the sample. Samples being processed for mitochondrial DNA testing only should be brought to 200 μl. Samples being processed for nuclear DNA only may be brought up to a lesser volume depending on the history of the case and/or the quality of the sample.
22. While it may initially seem unnecessary to cover the mortar and bone fragment with Parafilm, the reasons why will become abundantly clear upon striking the bone with the osteotome. The addition of the Parafilm prevents the bone fragments and splinters from departing the mortar.
23. Depending on how fresh the remains are, the bone powder will tend to be rather sticky. It may not pour out of the blender cup as easily as dried remains. It may be necessary to scrape the

powder from the blender cup using a scoopula (a spatula like scoop) or a small spatula.

24. The Mixer/Mill must be balanced and have two jars loaded at the same time. If pulverizing only one sample, place a jar containing only the ball in the other position.
25. The extraction buffer should be prewarmed to 56 °C. This can be achieved by placing the needed aliquots of buffer in the incubator shaker for approximately 45 min. This is an option when working with dried skeletal remains, but is recommended when working with fresh skeletal remains.
26. Unlike in the protocol for dried skeletal remains, the overnight incubation time is largely a necessity. An entire gram of bone powder will not sufficiently dissolve in an 8 h period.

Acknowledgements

The authors would like to thank all employees, past, present, and future, of the Armed Forces DNA Identification Laboratory (AFDIL), the American Registry of Pathology (ARP), and the Armed Forces Medical Examiner System (AFMES).

Disclaimer. The opinions expressed herein are those of the authors and not those of the U.S. Government, Department of the Navy, Department of the Army, the U.S. Army Medical and Development Command (MRDC), the Armed Forces Medical Examiner System (AFMES), the American Registry of Pathology (ARP), or the Armed Forces DNA Identification Laboratory (AFDIL). Mention of any product is merely a statement of use and should not be construed as an endorsement.

References

1. Loreille OM, Diegoli TM, Irwin JA et al (2007) High efficiency DNA extraction from bone by total demineralization. Forensic Sci Int Genet 1:191–195
2. Edson SM, Ross JP, Coble MD et al (2004) Naming the dead – confronting the realities of rapid identification of degraded skeletal remains. Forensic Sci Rev 16:63–90
3. Leney MD (2006) Sampling skeletal remains for ancient DNA (aDNA): a measure of success. Hist Archaeol 40:31–49
4. Mundorff AZ, Bartelink EJ, Mar-Cash E (2009) DNA preservation in skeletal elements from the World Trade Center disaster: recommendations for mass fatality management. J Forensic Sci 54:739–745
5. Huel R, Amory S, Bilić A et al (2012) DNA extraction from aged skeletal samples for STR typing by capillary electrophoresis. Methods Mol Biol 830:185–198
6. Gilbert MTP, Rudbeck L, Willerslev E et al (2005) Biochemical and physical correlates of DNA contamination in archeological human bones and teeth excavated at Matera, Italy. J Archaeol Sci 32:785–793
7. Higgins D, Kaidonis J, Townsend G et al (2013) Targeted sampling of cementum for recovery of nuclear DNA from human teeth and the impact of common decontamination measures. Investig Genet 4:18
8. Edson SM, Christensen AF (2013) Field contamination of skeletonized human remains with exogenous DNA. J For Sci 58:206–209
9. Adler CJ, Haak W, Donlon D et al (2011) Survival and recovery of DNA from ancient teeth and bones. J Archaeol Sci 38:956–964

Chapter 7

Extraction of DNA from Human Skeletal Material

Irena Zupanič Pajnič

Abstract

In recent years the recovery and analysis of DNA from skeletal remains has been applied to several contexts ranging from disaster victim identification to the identification of the victims of conflict. Here are described procedures for processing the bone and tooth samples including mechanical and chemical cleaning, cutting and powdering in the presence of liquid nitrogen, complete demineralization of bone and tooth powder, DNA extraction, DNA purification using magnetic beads, and the precautions and strategies implemented to avoid and detect contamination. It has proven highly successful in the analysis of bones and teeth from Second World War victims' skeletal remains that have been excavated from mass graves in Slovenia and is also suitable for genetic identification of relatively fresh human remains.

Key words Bone, Teeth, DNA extraction, Second World War victims, Identification, Contamination

1 Introduction

In cases where unidentified skeletonized human remains are found and identification cannot be performed using classical forensic methods, bones or teeth can be used for molecular genetic identification. In bones and teeth binding of DNA to hydroxyapatite aids its preservation [1]. However, DNA does degrade with time and the environmental conditions (temperature, humidity, pH, geochemical properties of the soil, and the presence of microorganisms) determine the level of molecular preservation [2–4]. The key factors for DNA preservation are ambient temperature and humidity in which the skeletal remains were located since the time of the organism's death until their exhumation and subsequent molecular genetic testing. Highly stable environments with little annual fluctuation in temperature or humidity are favorable for DNA preservation. The best examples of DNA preservation can be found in samples located in caves or permafrost, where low temperatures provide the best possible conditions for preservation. Warm, wet environments dramatically increase the degradation of DNA, resulting in extensive damage and fragmentation [5, 6].

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_7,

Another aspect that affects the quality and quantity of DNA in skeletal remains is the storage method used after their exhumation [7, 8]. The effectiveness of genetic typing is much higher with freshly exhumed skeletons rather than with skeletons that have been kept at room temperature for several years, particularly because of higher ambient temperatures and washing of the skeletal remains before storing, which can reduce the pH and salt content of samples. Freezing the skeletal remains is preferred in order to ensure the best preservation of the DNA [9]. According to Fulton [10] the most appropriate protocol for long-term storage of old specimens varies depending on how the specimens were collected. If a sample was frozen upon collection, it is ideal to maintain that temperature. If a sample was collected at room temperature, it should be stored in a cool, dry environment and may not benefit from being frozen, in particular if several freeze/thaw cycles are anticipated. In general, simply avoiding environmental conditions that are known to promote DNA damage is a key to sample preservation. A cool, dry, temperature-stable environment is ideal. Avoid heat, freeze/thaw cycles, and moisture [10].

The condition of the skeletal remains analyzed for forensic identifications is often not ideal for DNA recovery. In old bones and teeth, small amounts of degraded endogenous DNA, the presence of polymerase chain reaction (PCR) inhibitors, and the exceptional risk of contamination limit the success of DNA typing [11–13]. Hydrolytic and oxidative damage are likely to affect DNA over time. Oxidative damage results in modified bases, whereas hydrolytic damage results in deamination of bases and in depurination and depyrimidination. Both mechanisms reduce the number as well as the size of the fragments that can be amplified by PCR [14]. Extraction methods used for obtaining the DNA have to avoid overly aggressive treatments, such as high temperatures or use of strong detergents to reduce further degradation of the already damaged ancient DNA [15]. Failure to amplify DNA may also result from the presence of inhibitory low molecular weight compounds that co-extract with DNA and inhibit DNA polymerase in PCR. Contamination with modern DNA represents another major limitation to the molecular analysis of old bones because, as a result of its higher concentration and quality, contemporary DNA amplification is favored over that of the endogenous DNA in the sample [14].

Nuclear DNA is the preferred genome of amplification for forensic purposes as it is individually specific and provides biparental kinship information [16]. In the past, mitochondrial DNA testing was regularly employed in the identification of aged skeletal remains [17–19]. Recently, some researchers (among them is also our group) have reported the successful typing of nuclear STRs from old skeletal material [20–25]. We managed to obtain nuclear DNA for successful STR typing from skeletal remains excavated

from the Auersperg chapel archaeological site that were over 300 years old [26]. In addition, we successfully identified victims of massacres that took place during and after the Second World War in Slovenia [27, 28]. Positive identification of the victims of Second World War mass graves was possible where we could collect reference samples from living relatives. Skeletal remains were analyzed for the Konfin I mass grave, located in a karst cave, where 88 victims were killed, the karst cave Konfin II mass grave with 62 excavated skeletons, and four mass graves found in the Storžič forest (4 victims), Bodovlje gorge (25 victims), Mozelj (5 victims), and Mačkovec (16 victims) where massacre victims were excavated.

The extraction method should remove as many inhibitors as possible and should gain the maximum available DNA [29]. Decalcification with 0.5 M ethylene diamine tetra acetic acid—EDTA enables separation of bone cells from the bone mass [30]. When working with fresh bones and teeth, decalcification is not needed. This step is very important for old skeletal remains, because decalcification is crucial for gaining higher quantities of DNA [31]. Loreille and coworkers [32] succeeded in gaining a sufficient quantity of DNA when they used complete demineralization from old skeletal remain that gave no results without it. As shown by the latest studies, total demineralization is the best method of DNA extraction from old bone material [33, 34], since total demineralization significantly increases the proportion of full profiles, reflecting a correlation with better DNA quality. According to benefits of demineralization our protocol is based on total demineralization.

The method of DNA extraction was developed in our laboratory to acquire high-quality DNA from Second World War skeletal remains and from skeletal remains from archaeological sites. The same method is also used in our laboratory for molecular genetic identification of unknown decomposed bodies in routine forensic casework where only bones and teeth are suitable for DNA typing. We analyzed 111 bones and teeth from Second World War mass graves to evaluate this method [35] and additionally 54 Second World War skeletal remains samples and some 300 years old bones and teeth from archaeological sites for change extraction protocol from partial to total demineralization [26, 36]. We analyzed 111 bones and teeth from Second World War victims using a partial decalcification method and extracted up to 55 ng DNA/g from teeth, up to 100 ng DNA/g from femurs, and up to 30 ng DNA/g from tibias. The typing of autosomal and Y-STR loci was successful in 95 % of the bones and teeth (there were approximately 20 % of partial profiles) and mtDNA in 96 % (HVSI) to 98 % (HVSII) of the samples analyzed [35]. Extracting genomic DNA using the total demineralization method from 54 Second World War skeletal remains samples gave us almost complete autosomal STR profiles in 52 out of the 54 samples; very few allelic drop-outs were observed in comparison to STR typing of extracts obtained with

partial demineralization [35, 36]. We successfully identified more victims of massacres that took place after the Second World War in Slovenia [27, 28]. We also performed an efficiency studies of different commercially available quantification and amplification kits for autosomal STR typing of skeletal remains excavated from the Second World War mass graves and concluded that they are highly reliable for human quantification and STR typing of old bones and teeth with the DNA extraction method optimized in our laboratory [36–38]. Amplification efficiency can be improved by addition of BSA (final concentration 40 ng/μl), which has the ability to bind to enzyme inhibitors present in DNA extract, and by increasing the number of amplification cycles. We used BSA in some bone and tooth extracts to overcome the PCR inhibition and we used three additional amplification cycles only in low-template bone and teeth DNA samples for STR DNA typing with commercially available amplification kits [28, 38]. When testing the performance of amplification kits with the extended European Standard Set (ESS) of loci on a sample of 102 seventy years old bones and teeth, DNA typing was successful in almost all of the samples. The method of DNA extraction described here has proved to be highly efficient because we obtained up to 131 ng DNA/g of bone and for the most of Second World War samples complete genetic profiles of autosomal STRs [37].

When drying the bones and teeth at 50 °C in an oven or incubator, the protocol yields DNA extracts in 28 h; otherwise two more days are needed to complete the extraction procedure when bone and tooth samples are dried at room temperature. The method proved effective from relatively small amount of bone or tooth powder (0.5 g). Less than 0.5 g can be used in case of small sample. In that case the volume of EDTA solution should be adjusted proportionally. The purification procedure using a Biorobot EZ1 (Qiagen) device doesn't use any aggressive organic solvents like phenol or chloroform. It is automated and takes only 20 min to complete. It is based on technology of magnetic particles that are covered with silicon. Such magnetic particles are very efficient for binding DNA, especially in the presence of chaotropic salts (like guanidine thiocyanate (GuSCN) or guanidine hydrochloride (GuHCl)). These salts are highly efficient for nucleic acid cleaning [39]. Chaotropic salts lyse cells, denature proteins, inactivate nucleases, and accelerate the binding of DNA to the paramagnetic particles covered with silicon. The whole extraction process is done in a huge filter tip that is thrown away after the procedure is finished. The rest of the extraction reagents are safely placed in a container—cartridge for single use only. Therefore, no manual pipetting is needed. This is very important for prevention of contamination. High efficiency of magnetic particles in DNA extraction was confirmed in several studies [40–43]. The purification using magnetic particles can be adapted also

to other robotic machines or can be performed manually using manual kits with magnetic particles from different suppliers. We successfully purified DNA from old skeletal material using also an AutoMate Express Instrument—Applied Biosystems [44].

Any validated human quantification and multiplex STR kit may be used with this extraction method. For autosomal STR typing of skeletal remains excavated from Second World War mass graves in Slovenia especially new amplification kits with the extended ESS loci proved highly reliable [36, 37]. Depending on the reference samples, different multiplex kits (autosomal STRs and Y-chromosomal STRs) and also sequencing of mtDNA may be needed to provide enough genetic markers for sufficiently significant results from genetic kinship analyses.

2 Materials

2.1 Chemicals

1. Sodium hypochlorite—bleach (Kemika); 6 % solution.
2. Alconox detergent (Sigma-Aldrich); 5 % solution.
3. Bi-distilled water (Sartorius-Stedim Biotech or Millipore).
4. Liquid nitrogen.
5. Ethanol (Merck); 80 %.
6. Ethylene diamine tetra acetic acid—EDTA (Promega); 0.5 M solution pH 8.0.
7. EZ 1 DNA Investigator Kit (Qiagen) contains:
   (a) Buffer G2
   (b) Proteinase K
8. DTT (Sigma-Aldrich); 1 M solution.
9. cRNA (EZ 1 DNA Investigator Kit, Qiagen).
10. Ultrapure distilled water HPLC grade (Gibco).
11. DNA Away™ (Molecular BioProducts).
12. HCl (Merck); 1 M solution.
13. Sodium acetate (Merck); 2 M solution pH 5.2.
14. NaOH (Merck); 5 M solution.

2.2 Consumable Goods

1. Filter tips.
2. pH indicator strips.
3. 50-ml Falcon tubes (Sarstedt).
4. 1.5- or 2-ml tubes (Eppendorf).
5. Scalpel blade.
6. Rough part of a dish sponge.
7. Plastic vial.

8. Cellulose (thin layer).
9. Sterile latex gloves.
10. EZ 1 DNA Investigator Kit (Qiagen) contains:
    (a) Cartridges with purification reagents
    (b) Sample tubes
    (c) Elution tubes
    (d) Filter tips with tip holders

2.3 Equipment

- Saw (Aesculap).
- Sterile saw blades (Aesculap).
- Standard laboratory equipment such as freezer and refrigerator for storing extracts and chemicals.
- Microbiological safety cabinet MC 3 (Iskra Pio).
- Laminar flow hoods MC 1 (Iskra Pio).
- Holding vice (Proxxon).
- High-speed grinding machine (Dremel).
- Circular diamond saws (Proxxon).
- Drilling bits (Proxxon).
- Shaker.
- Balance (A&D Company).
- Metal grinding vials (25 ml) with metal balls (2r = 20 mm) (Tehtnica—Domel).
- Bead Beater MillMix 20 (Tehtnica—Domel).
- Tweezers.
- Forceps.
- Mortuary needle.
- Hammer.
- Spatulas.
- Pipettes.
- Vortexes.
- Thermomixer comfort (Eppendorf) or any other shaker with temperature setting, suitable for use with 50-ml tubes.
- Centrifuge Megafuge 1.0 (Heraeus) or any other centrifuge suitable for use with 50-ml tubes.
- Instrument BioRobot EZ 1 (Qiagen).
- EZ 1 DNA Investigator Card (Qiagen).

3 Methods

3.1 Measures for Preventing and Detecting DNA Contamination

In the process of DNA typing, we encounter not only highly degraded DNA but also very small amounts of endogenous DNA which are very susceptible to contamination with modern DNA and are difficult to differentiate from the far more common modern (exogenous) DNA. Unfortunately, contamination is a serious

problem in investigations of DNA obtained from old skeletal remains [45–49]. Many chemical and physical environmental factors can have influence on contamination of skeletal remains with DNA from bacteria and fungi. That kind of contamination is not possible to prevent; however, it can affect the success of mitochondrial DNA and nuclear DNA typing. We are able to implement procedures that minimize the possibility of contamination with modern human DNA. Contamination of the endogenous DNA of bones and teeth with modern DNA can occur during exhumation, improper storage of the skeletal remains, and anthropological investigations [50]. Surface contamination can often occur due to improper handling of skeletal remains with bare hands. Contamination can also occur in the molecular genetic laboratory during the process of DNA typing where contaminating DNA can be located on the laboratory plastics and reagents. DNA fragments can also be present in the air in aerosol particles [51]. Therefore, it is necessary to consider the recommendations to prevent contamination. An elimination database, containing the profiles of the individuals that participated in the exhumation and subsequent analysis of the remains, can be used to check for authenticity of genetic profiles obtained from old skeletal remains and allows traceability in the case of contamination. Extraction negative controls have to be included in every extraction batch and PCR-negative controls in every amplification reaction to verify the purity of the extraction and amplification reagents and plastics. Doing so allows us to trace contamination in the event of its occurrence [52]. According to Rohland and Hofreiter [15] when processing more than seven bone or tooth samples, two or more extraction controls should be included in analyses. At least two samples must be typed for each skeleton, and it is necessary to obtain identical genetic profiles from both of them [52]. Physical separation of pre- and post-PCR areas is important to minimize the possibility of contamination through previously amplified products.

We eliminate surface contamination through different methods. The most important are washing in bi-distilled water, detergent, and ethanol; radiation with UV light and removing the bone surface and acquiring the bone or tooth material directly from the inside of the specimen. For successful decontamination, we usually use a combination of all listed methods. At the Laboratory of Molecular Genetics in the Institute of Forensic Medicine in Ljubljana, we follow stringent recommendations for prevention of contamination [1, 2, 52–62]. We use the following measurements to prevent contamination in the laboratory:

1. To prevent contamination with our own biological material, always use clean, sterile gloves (use double laboratory gloves) and change for every new sample. Use disposable surgical masks, caps, shoe covers, and disposable laboratory coats.

2. Clean the entire working surface before and after any work is performed with bleach (6 % sodium hypochlorite) followed by sterile bi-distilled water and 80 % ethanol, and/or nightly UV irradiation. Surfaces are cleaned in the same manner between each set of skeletal remains.
3. Clean all tools for cleaning, abrasion, and grinding of bones and teeth after use with bleach (6 % sodium hypochlorite) or with DNA Away™—Molecular BioProducts, which is very efficient in decontamination. Tools have to be exposed to the bleach for 15–30 min. Wash away the detergent with several washes with sterile bi-distilled water and 80 % ethanol and leave tools to air dry. Finally, all tools are sterilized using incubation at 134 °C for 45 min in a Tecno-Gaz sterilizer, and UV irradiated with shortwave (254 nm) UV source at least overnight or up to 72 h and for 30 min directly before starting to work.
4. Put all the reagents, tools, and laboratory plastics after sterilization under UV light at least overnight or up to 72 h. We expose all the listed material to UV light for 30 min directly before starting to work.
5. To avoid cross-contamination among samples a different set of equipment is used for each sample (such as grinding vials, cutting saw blades, drilling bits, tweezers, forceps, mortuary needles, and spatulas). These are cleaned and then stored in a way that makes it accessible but minimizes the possibility of cross-contamination with dust from cutting/grinding.
6. Take clean tools for each bone or tooth specimen.
7. Analyze bone and teeth samples separately from reference samples (e.g., for the elimination database). We use physically separated room for processing bone and tooth samples. Analyses of skeletons should be also temporally separated from reference samples and elimination database samples.
8. The separation of pre- and post-PCR procedures must be provided to prevent contamination with previously amplified products (all the equipment and protective clothing from post-PCR room never entered the pre-PCR laboratory). Amplified products from the post-PCR room should never be introduced to pre-PCR laboratory.
9. It is necessary to separate the dust-producing working steps from the contamination-susceptible steps like buffer preparation and PCR setup. We have different rooms in pre-PCR laboratory to separate each step in the bone typing procedure. We have room for cleaning and grinding the bones and teeth. In that room we clean the bones mechanically in a closed microbiological safety cabinet MC 3 (Iskra Pio) to capture and remove the bone powder that is released into the air during drilling and cutting. It has strong airflow to the filters that col-

lect the dust at the bottom of the chamber. The second room is used for preparation of buffers and solutions. The extraction room is used for decalcification, extraction, and purification, and the PCR room is used for the setup of PCR reagent mix (first hood) and addition of DNA extracts to the PCR (second hood). In each room we have laminar flow hoods with short-wave (254 nm) UV source and hepa filters. The laboratory setup must prevent dust from contaminating the rest of the process in DNA typing of skeletal remains.

10. Pre-PCR laboratory undergoes regular decontamination (washing with bleach, water, and ethanol). After the work the laminar flow hoods are irradiated at least overnight and for 30 min directly before starting to work.
11. To detect any possible contamination with DNA or previously amplified PCR products of reagents or laboratory plastics, we always use at least one negative PCR control.
12. For monitoring the cleanliness of the isolation reagents and laboratory plastics, and cross-contamination during the procedure we always use isolation negative control.
13. All genetic profiles obtained from skeletal remains are compared to elimination database.
14. Always use filter tips to minimize the risk of cross-contamination owing to DNA aerosols. Tips are exposed to UV light before use. Always use separate pipette tip for each sample to avoid cross-contamination.
15. We use the room for cleaning, grinding, decalcification, and extraction of DNA from bones and teeth exclusively for this kind of biological material and not for any other sample that contains high-template DNA (saliva, blood samples).
16. We isolate DNA from bones and teeth at least twice (from a different skeletal element of the same individual when possible) to check the results of genotyping and for interpretation reproducible results are used.

3.2 Reagent Preparation

All solutions should be prepared using HPLC-grade ultrapure water (Gibco) that was beforehand UV illuminated at least overnight.

3.2.1 Preparation of 5 % Alconox

Weigh 25 g of detergent Alconox (Sigma-Aldrich) on the balance into the 50-ml Falcon tube. Put it into 500-ml bottle, and add 500 ml ultrapure distilled water (Gibco). Put the bottle into hot water to dissolve the detergent. Expose the bottle to UV irradiation at least overnight or up to 72 h. Store at room temperature. The solution is stable for several months.

3.2.2 Preparation of 80 % Ethanol

Pour off 400 ml of absolute ethanol (Merck) into 500-ml bottle and add 100 ml ultrapure distilled water (Gibco). Expose the

bottle to UV irradiation at least overnight or up to 72 h. Store at room temperature. The solution is stable for several months.

3.2.3 Preparation of 0.5 M Ethylene Diamine Tetra Acetic Acid (EDTA) (pH 8.0)

Weigh 46.53 g EDTA (Promega) on the balance into 50-ml Falcon tube. Put it into 250- or 500-ml bottle and fill to 200 ml with ultrapure distilled water (Gibco). Adjust pH to 8.0 with 5 M NaOH. The final volume is 250 ml. Fill with ultrapure distilled water (Gibco) to 250 ml. Autoclave. Expose to UV irradiation at least overnight or up to 72 h. Store at room temperature. The solution is stable for several months.

To prepare 5 M NaOH: weigh 20 g NaOH (Merck) into a bottle and fill to 100 ml with ultrapure distilled water (Gibco) and autoclave. Store at room temperature. The solution is stable for several months.

3.2.4 Preparation of 1 μg/μl cRNA

Add 310 μl of ultrapure distilled water (Gibco) to lyophilized cRNA (included in the EZ 1 DNA Investigator Kit, Qiagen) and vortex at maximal speed for 10 s. Make aliquots—pipette 10 μl of the dissolved cRNA into 200-μl tubes. Deep freeze at −20 °C. Stored in the freezer it will remain stable for 6 months. Just before use take the frozen cRNA from the freezer and add it to the sample; discard the rest.

3.2.5 Preparation of 1 M DTT

Weigh 154 mg DTT (Sigma-Aldrich) in a 2-ml Eppendorf tube; add 1 ml ultrapure distilled water (Gibco) and 5 μl 2 M sodium acetate pH 5.2. Vortex, aliquot, and freeze at −20 °C. Just before use take the frozen 1 M DTT from the freezer. Stored in the freezer it will remain stable for 6 months.

To prepare 2 M sodium acetate: weigh 8.203 g anhydrides CH_3COONa (Merck) into 50-ml Falcon tube and fill to 50 ml with ultrapure distilled water (Gibco). Calibrate to pH 5.2 with 1 M HCl. Autoclave and store at room temperature. The solution is stable for several months.

To prepare 1 M HCl: put 10 ml concentrated HCl (Merck) into an autoclaved bottle and fill to 100 ml with ultrapure distilled water (Gibco). Don't autoclave. Store at room temperature. The solution is stable for several months.

3.3 Bone and Tooth Sample Preparation

3.3.1 Bone and Tooth Sample Selection

Long bones and teeth are the most appropriate samples for molecular genetic testing, as the DNA in them can stay well preserved for a long time [63–65]. The skull bones are the least suitable for genetic investigation according to Edson et al. [65]. From the comparative study of the performance of nuclear DNA typing of skeletal remains (we typed teeth, femurs, and tibiae but we didn't type any smaller elements of the hands and feet) from the mass graves of the Second World War, our laboratory discovered that teeth are the most suitable for typing, followed by the femur bones and tibiae [36, 37]. Similar conclusions were also reached by Miloš

et al. [63] and Misner and colleagues [64]. According to the experience of Keyser-Tracqui and Ludes [14], heavy (dense) bone is better than more brittle bone, which has lost lipid and collagen and has therefore increased porosity. Long bones (femur, tibia, and humerus) are preferred over rib or other thin bones and compact (cortical) bone is preferred to spongy bone. Rohland and Hofreiter [15] recommend getting the bone powder from diaphysis of long bones and use of dentine rather than enamel from teeth, as dentine is assumed to contain more DNA. Recently Mundorff et al. [66] and Mundorff and Davoren [67] found that smaller elements of the hands and feet (metatarsals, metacarpals, phalanges) were very similar or even better in DNA yield as both femora and tibiae. These bones can be easily sampled with a disposable scalpel, and thus reduce potential DNA contamination. Based on recent studies the current recommendations for preferential testing of long bones from the legs may need to be reevaluated and the sampling strategy for laboratories typing bone samples may change in the future.

The amount of DNA from individual and between groups of teeth varies strongly; quality and quantity of isolated DNA also depend on the tooth pathology, previous dental procedures, elapsed time since the extraction of the tooth until the isolation of DNA, and the donor's age [68]. Maximum DNA is obtained from the whole pulverized tooth, making it possible to capture DNA located in the hard dental tissues [69]. The amount of DNA depends on the size of the dental pulp and type of teeth; the molars are the richest source of DNA. Teeth suitable for DNA isolation appear in the following order: endodontically untreated molar, premolar, canine, and incisor and endodontically treated molar, premolar, canine, and incisor. At our laboratory, we select for genetic testing one long bone (preferably femur) and two teeth (preferably well-preserved and endodontically untreated molars) from each individual skeleton found in the Second World War graves; that is only possible through the excavation of skeletons in anatomic position. If excavation of skeletons was not carried out in the anatomical position, we select for molecular genetic investigations all left or all right femurs found in the grave. All skeletal material is photo-documented, appropriately labeled, and fragments of bones and teeth sampled for molecular genetic analyses are frozen at –20 °C until the DNA isolation procedure.

3.3.2 Bone and Tooth Sample Cleaning

Research has shown that washing and improper handling of remains contaminate their surface and can even penetrate into deeper layers. This kind of inner contamination is dependent on the stage of porosity and preservation of the remains [70–72]. Therefore, skeletal remains must be cleaned mechanically and chemically and teeth must be UV irradiated instead of mechanical cleaning. Although no procedures are 100 % efficient when removing contamination introduced during excavation, storage or

collection exists, cleaning improves the ratio between endogenous and contaminating DNA and it may reduce the amount of inhibitors introduced into the extraction [15].

We clean skeletal remains in a closed microbiological safety cabinet MC 3 (Iskra Pio) in a room designed exclusively for processing old skeletal remains. Bone samples are cleaned mechanically (physical removal of the surface using a rotary sanding tool (Dremel)) and chemically (washing in detergent, water, and ethanol), while tooth samples are cleaned chemically (washing in detergent, water, and ethanol) and irradiated with UV light for 2×30 min with the tooth rotated 180° between each exposure prior to grinding into a powder. To prevent bone warming during drilling and cutting, we frequently use liquid nitrogen to cool the bone and we use lower speed setting for abrasion and cutting. Warming of bone may cause degradation of endogenous DNA [59].

Between each sample all tools for drilling, cutting, and grinding of bones are cleaned by washing with water, bleach (6 % sodium hypochlorite) or DNA Away™ (Molecular BioProducts), sterile bi-distilled water (Sartorius-Stedim Biotech or Millipore), and 80 % ethanol. Finally, everything is sterilized and UV irradiated at least overnight or up to 72 h. The steps for cleaning the bone and tooth samples are as follows:

1. With the use of a sterile clean saw blade cut 8–10 cm long and 2–3 cm wide fragment of femur just below trochanter (the same size can be applied for the rest long bones, too). Remove the molar with forceps from upper or lower jawbone. Store the bone fragment or tooth sample in the labeled 50-ml Falcon tube. If you don't work on that piece of bone or tooth sample immediately, freeze it at –20 °C for a long-term storage (*see* **Note 1**). The saw blade and the forceps must be changed for every new bone or tooth.
2. Remove dirt, soil, and any other material from the surface of the bone or tooth with a sterile scalpel blade, strong spatula, or needle (e.g., Mortuary needle) (*see* **Note 2**).
3. Put the bone or tooth into a plastic vial and with the use of a rough part of a sterilized and UV irradiated dish sponge wash it in sterile bi-distilled water (Sartorius-Stedim Biotech or Millipore) with added mild detergent (add few ml of 5 % Alconox detergent to the water). Put the bone or tooth into 50-ml Falcon tube and wash it three times with sterile bi-distilled water (Sartorius-Stedim Biotech or Millipore).
4. Dry the bone fragment or the tooth overnight; cover it with thin layer of cellulose paper (*see* **Note 3**).
5. Put everything you need for bone or teeth processing next day under UV light.

6. The next day weight bone or tooth sample and remove surface contamination with polishing (for teeth we use only UV irradiation on both side for 30 min). For removing surface contamination from the bone sample, use a closed microbiological safety cabinet. Fasten the dried bone into holding vice. With the high-speed grinding machine and drilling bit sand down surface 1–3 mm layer of the bone. Remove the surface layer from inner side of the bone fragment. Use liquid nitrogen to cool down the bone fragment and repeat the drilling on outer side of the bone.
7. Put mechanically cleaned bone into liquid nitrogen to cool.
8. Fasten the bone into holding vice. With the use of a circular diamond saw make notches in the shape of net (make small squares in dimension of 5×5 mm) on the outer surface of the bone. Use liquid nitrogen to cool down the bone fragment and make the same net also on the inner side of the bone (both sides have to look like a net).
9. Cut cleaned part of the bone with the circular diamond saw from the rest of the bone and put it into a sterile 50-ml Falcon tube using sterile tweezers. **Steps 6–9** are performed in a fume hood to control dust. The weight of cleaned piece of bone is approximately 2–3 g. The rest of the bone sample store for additional extraction if needed. The next step is chemical cleaning of bones and teeth with 5 % Alconox detergent (Sigma-Aldrich), sterile bi-distilled water (Sartorius-Stedim Biotech or Millipore), and 80 % ethanol.
10. Wash out the bone or tooth sample for 1 min with bi-distilled water (Sartorius-Stedim Biotech or Millipore). Mildly shake it.
11. Pour off water from the Falcon tube.
12. Wash the bone or tooth sample with 5 % Alconox detergent (Sigma-Aldrich) and mildly shake it for 15–30 s on a shaker. Pour off the detergent from the Falcon tube (*see* **Note 4**). To remove detergent four steps of washing follow.
13. Wash the bone or tooth sample with bi-distilled water (Sartorius-Stedim Biotech or Millipore) for 3 min and mildly shake it.
14. Pour off water from the Falcon tube.
15. Repeat washing with water once again.
16. Wash the bone or tooth sample in 80 % ethanol for 30 s and mildly shake it.
17. Pour off ethanol from the Falcon tube.
18. Repeat washing with ethanol once again.
19. Dry clean fragment of the bone or tooth sample overnight and cover it with thin layer of cellulose paper (*see* **Note 3**).
20. All listed steps are done to remove possible contaminants from previous handling from the sample surface.

21. Clean the entire working surface of a closed microbiological safety cabinet after the work or between working with different skeletal remains with bleach (6 % sodium hypochlorite), sterile bi-distilled water (Sartorius-Stedim Biotech or Millipore), and 80 % ethanol using paper towels.
22. Put everything you need for your work next day under UV light.
23. Next day put the dried bone or tooth sample into 50-ml Falcon tube and weigh it on the balance and write down the results on Falcon tube.

3.3.3 Bone and Tooth Sample Powdering

It is necessary to obtain very fine bone or tooth powder to extract enough quantity of DNA from old skeletal remains. Demineralization is better and faster with very small pieces of powder, so generate as fine powder as possible to maximize the surface area of the sample that will eventually contact the chelation solution [15]. We use homogenizer Bead Beater MillMix 20 (Tehtnica—Domel) and liquid nitrogen to obtain fine bone and tooth powder. We cool metal vials and bone or tooth samples in liquid nitrogen to avoid overheating during powdering and then we grind them for 1–2 min at 30 Hz. The powder we get is transferred into a sterile 50-ml Falcon tube. We weight the powder and use it in the extraction process. We pulverize skeletal remains in a room designed exclusively for processing old skeletal remains. Grinding vials need to be thoroughly cleaned before reuse. Between each sample, they have to be cleaned by washing with water, bleach (6 % sodium hypochlorite) or DNA Away™ (Molecular BioProducts), sterile bi-distilled water (Sartorius-Stedim Biotech or Millipore), and 80 % ethanol. Finally, the grinding vials are sterilized and UV irradiated at least overnight or up to 72 h and additionally 30 min before use. A sufficient number of grinding vials are required to prepare more than one sample per day. The steps for powdering the bone and tooth samples are as follows:

1. Pour liquid nitrogen over the bone or tooth sample in Falcon tube and wait until it evaporates.
2. Pour liquid nitrogen into the lower part of the metal vial and after evaporation repeat twice to cool the grinding vial.
3. With help of tweezers move the bone or tooth sample into the sterile and UV irradiated glove finger, wrap it in several sterile cellulose papers, and break it into smaller pieces with hammer (bone will break on the notches previously made with circular diamond saw).
4. Move the broken pieces in dimension of 5×5 mm with tweezers into the cooled lower part of the metal vial. Add metal ball and close the vial with its own metal lid. Grind to obtain bone powder at 30 Hz 1–2 min in the Bead Beater MillMix 20 (Tehtnica—Domel).

5. Move the bone or tooth powder into the sterile 50-ml Falcon tube and weigh it on the balance. Write the result on the Falcon tube.
6. The sample powder can be stored at 4 °C while in use, but should be subjected to the extraction as soon as possible. For long-term storage store it at –20 °C.

3.3.4 Bone and Tooth Sample Decalcification

Genomic DNA is obtained from 0.5 g of bone or tooth powder incubated in 10 ml of 0.5 M ethylene diamine tetra acetic acid—EDTA pH 8 overnight at 37 °C for decalcification. EDTA is a strong chelator that is able to bind metallic ions such as calcium in the bone or tooth powder and allows for its removal. High amounts of EDTA are necessary to dissolve part of the hydroxyapatite matrix specific to bone and teeth samples [15]. For total demineralization 15 ml of 0.5 M EDTA per g of bone or tooth powder is needed. That amount of EDTA can theoretically bind only the amount of calcium contained in 1 g of bone or tooth powder [32]. At the end of decalcification process, the precipitate is washed with sterile bi-distilled water (Sartorius-Stedim Biotech or Millipore). The steps for decalcification of bone or tooth samples are as follows:

1. Weigh 0.5 g of bone or tooth powder in the 50-ml Falcon tube. Use sterile spatula.
2. Prepare another 50-ml Falcon tube used for isolation blind control. Put reagents for DNA isolation in this tube only (do not put bone or tooth powder in this tube). Blind control has to be treated identically to the experimental samples throughout the procedure.
3. Add 10 ml of 0.5 M EDTA and vortex for 10 s (*see* **Note 5**).
4. Incubate bone or tooth powder and extraction negative control overnight at 37 °C and mix at 750 rpm on the Thermomixer comfort (Eppendorf).
5. Put the necessary material for working on the bone or tooth next day under UV light.
6. Centrifuge at 1300 × *g* for 15 min, a pellet of residual powder that is typically seen at this point.
7. Pipette and discard whole supernatant. In extraction negative control leave only approximately 100 µl of supernatant.
8. Add 10 ml of sterile bi-distilled water (Sartorius-Stedim Biotech or Millipore) and vortex at high speed for 10 s. Add 10 ml of sterile bi-distilled water (Sartorius-Stedim Biotech or Millipore) also to the extraction blind control sample.
9. Centrifuge at 1300 × *g* for 15 min.
10. Pipette and discard whole supernatant (for the blind control leave only approximately 100 µl of supernatant).

3.4 DNA Extraction and Purification

We don't use the organic extraction with phenol/chloroform/isoamyl alcohol for extraction and purification of DNA. In our method the DNA is purified in a Biorobot EZ1 (Qiagen) device using the EZ1 DNA Investigator Card and EZ1 DNA Investigator Kit (Qiagen) (*see* Chapter 5). Both phenol and chloroform are dangerous and treatment with them should always be performed in a vented fume hood. Because of their toxicity, it is much safer to use other efficient methods of purification of DNA. We optimized extraction and purification of DNA in Biorobot EZ1 device (Qiagen) to obtain sufficient amount of bone and tooth DNA for successful STR typing.

3.4.1 Extraction of DNA

An extraction buffer, proteinase K, and DTT are added to the precipitate and incubated for 2–3 h at 56 °C. Higher incubation temperature improves the digestion of the bone or tooth powder and thereby releases more DNA, especially in cases when the powder used is relatively coarse (*see* **Note 6**). Proteinase K is an endolytic serine protease that cleaves proteins, reducing them to their constituent amino acids. DTT is a reducing agent that can cleave cysteine-cystine bridges and disrupt the tertiary structure of some proteins and allow further degradation. The steps for extraction of DNA from bone or tooth samples are as follows:

1. Add 100 µl G2 buffer (EZ 1 DNA Investigator Kit, Qiagen) to the pellet and to the blind control.
2. Add 60 µl Proteinase K (EZ 1 DNA Investigator Kit, Qiagen) to the pellet and to the blind control (*see* **Note 7**).
3. Add 20 µl 1 M DTT to the pellet and to the blind control (*see* **Note 8**).
4. Vortex for 10 s (*see* **Note 5**).
5. Incubate at 56 °C and 750 rpm in the Thermomixer comfort (Eppendorf) for 2–3 h.

3.4.2 Purification of DNA

After centrifugation, the supernatant is taken to purify the DNA in a Biorobot EZ1 device (Qiagen). The DNA extract should be colorless. We store the DNA that was extracted until forthcoming steps of quantification and genotyping of nuclear and mtDNA at 4 °C while in use and at −20 °C for long-term storage. DNA is susceptible to damage from repeat freeze-thaw cycles and should be defrosted as infrequently as possible [73]. We always include negative controls in the process of extraction to check cleanliness of laboratory plastics and reagents. The steps for purification of DNA from bone or tooth samples are as follows:

1. Centrifuge the bone or tooth lysate and the blind control sample at $600 \times g$ for 2 min.
2. Pipette 200 µl (for the blind control sample) and 400 µl (for the bone or tooth lysate) of the supernatant into sample tube

(Qiagen). Keep the remaining supernatant (up to 300 μl); you may wish to retain it for a second round of extraction with Biorobot EZ1 device. Store the remaining lysate in a 1.5- or 2-ml Eppendorf tube and freeze it at –20 °C (*see* **Note 9**).

3. Add 1 μl cRNA—Qiagen (concentration of 1 μg/μl) (*see* **Note 8**).
4. Put the sample tubes into Biorobot EZ1 instrument (Qiagen).
5. Choose: trace protocol, water elution, and 50 μl volume of elution in the software menu of Investigator software Card (Qiagen).
6. After the automated purification procedure, you get 50 μl isolated DNA sample. Close elution tubes with extracted DNA and blind control sample and store them at 4 °C while in use and then place it at –20 °C for long-term storage (*see* **Note 10**).

4 Notes

1. According to Rohland and Hofreiter [15], the sample can be stored at room temperature, but we prefer to be cautious and store at –20 °C.
2. This step is important because dirt may introduce a variety of inhibitory substances to the extraction procedure, and therefore to the extract itself; these substances may interfere or even completely block subsequent enzymatic manipulations of the DNA extracts [15].
3. It is also possible to dry bones and teeth at 50 °C for 2 h in an oven or incubator.
4. Any detergent carryover will degrade the DNA and reagents in subsequent steps of the DNA extraction; thus it is extremely important that detergent is removed completely.
5. Vortex at slow speed to prevent sticking of pellet on the walls of Falcon tube.
6. Be aware that higher temperature may cause further damage or degradation of the DNA [15].
7. In our protocol we use more proteinase K than it is recommended by the manufacturer and supplied with EZ 1 DNA Investigator Kit (Qiagen). Accordingly some extra proteinase K (Qiagen) has to be ordered.
8. Thaw it immediately before use.
9. The second extraction will usually contain lower amount of DNA compared to the first and you can use the first one for nuclear DNA typing and the second one for mtDNA typing.
10. It may be useful to subdivide the final extract into aliquots of 10 μl and to use these as necessary (successive freeze-thaw cycles can damage the DNA over time).

Acknowledgements

The author gratefully acknowledges the contribution of Barbara Gornjak Pogorelc and Katja Vodopivec Mohorčič towards processing and DNA typing of bones and teeth.

References

1. Kemp BM, Smith DG (2005) Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. Forensic Sci Int 154:53–61
2. Paabo S (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci U S A 86:1939–1943
3. Höss M, Jaruga P, Zastawny TH, Dizdaroglu M, Pääbo S (1996) DNA damage and DMA sequence retrieval from ancient tissues. Nucleic Acids Res 24:1304–1307
4. Poinar HN, Höss M, Bada JL, Pääbo S (1996) Amino acid racemization and the preservation of ancient DNA. Science 272:864–866
5. Smith CI, Chamberlain AT, Riley MS, Cooper A, Stringer CB, Collins MJ (2001) Neanderthal DNA: not just old but old and cold? Nature 410:771–772
6. Smith CI, Chamberlain AT, Riley MS, Stringer C, Collins MJ (2003) The thermal history of human fossils and the likelihood of successful DNA amplification. J Hum Evol 45:203–217
7. Burger J, Hummel S, Herrmann B, Henke W (1999) DNA preservation: a microsatellite-DNA study on ancient skeletal remains. Electrophoresis 20:1722–1728
8. Pruvost M, Schwarz R, Correia VB, Champlot S, Braguier S, Morel N et al (2007) Freshly excavated fossil bones are best for amplification of ancient DNA. Proc Natl Acad Sci U S A 104:739–744
9. Malmstrom H (2007) Ancient DNA as a means to investigate the European neolithic. PhD Thesis. Uppsala, Uppsala University
10. Fulton LT (2012) Setting up an ancient DNA laboratory. In: Shapiro B, Hofreiter M, editors. Ancient DNA - Methods and Protocols. New York: Humana Press Inc 1–11
11. Alaeddini R, Walsh SJ, Abbas A (2010) Forensic implications of genetic analyses from degraded DNA-a review. Forensic Sci Int Genet 4:148–157
12. Alaeddini R (2012) Forensic implications of PCR inhibition - a review. Forensic Sci Int Genet 6:297–305
13. Lee HY, Park MJ, Kim NY, Sim JE, Yang WI, Shin KJ (2010) Simple and highly effective DNA extraction methods from old skeletal remains using silica columns. Forensic Sci Int Genet 4:275–280
14. Keyser-Tracqui C, Ludes B (2005) Methods for the study of ancient DNA. Methods Mol Biol 297:253–264
15. Rohland N, Hofreiter M (2007) Ancient DNA extraction from bones and teeth. Nat Protoc 2:1756–1762
16. Lee EJ, Luedtke JG, Allison JL, Arber CE, Merriwether DA, Steadman DW (2010) The effects of different maceration techniques on nuclear DNA amplification using human bone. J Forensic Sci 55:1032–1038
17. Anslinger K, Weichhold G, Keil W, Bayer B, Eisenmenger W (2001) Identification of the skeletal remains of Martin Bormann by mtDNA analysis. Int J Legal Med 114:194–196
18. Stone AC, Starrs JE, Stoneking M (2001) Mitochondrial DNA analysis of the presumptive remains of Jesse James. J Forensic Sci 46:173–176
19. Palo JU, Hedman M, Söderholm N, Sajantila A (2007) Repatriation and identification of Finnish World War II soldiers. Croat Med J 48:528–535
20. Irwin JA, Edson SM, Loreille O, Just RS, Barritt SM, Lee DA et al (2007) DNA identification of "earthquake McGoon" 50 years postmortem. J Forensic Sci 52:1115–1118
21. Irwin JA, Leney MD, Loreille O, Barritt SM, Christensen AF, Holland TD et al (2007) Application of low copy number STR typing to the identification of aged, degraded skeletal remains. J Forensic Sci 52:1322–1327
22. Lee HY, Kim NY, Park MJ, Sim JE, Yang WI, Shin KJ (2010) DNA typing for the identification of old skeletal remains from Korean War victims. J Forensic Sci 55:1422–1429
23. Vanek D, Saskova L, Koch H (2009) Kinship and Y-chromosome analysis of 7th century human remains: novel DNA extraction and typing procedure for ancient material. Croat Med J 50:286–295
24. Bogdanowicz W, Allen M, Branicki W, Lembring M, Gajewska M, Kupiec T (2009) Genetic identification of putative remains of

the famous astronomer Nicolaus Copernicus. Proc Natl Acad Sci U S A 106:12279–12282

25. Zupanič-Pajnič I (2013) Genetic identification of Second World War victim's skeletal remains. Lap Lambert Academic Publishing, Saarbrucken
26. Zupanič-Pajnič I (2013) Molecular genetic analyses of 300-year old skeletons from Auersperg tomb. Zdravniski Vestnik 82:796–808
27. Zupanič-Pajnič I (2008) Molecular genetic identification of the Slovene home guard victims. Zdravniski Vestnik 77:745–750
28. Zupanič-Pajnič I, Pogorelc BG, Balazic J (2010) Molecular genetic identification of skeletal remains from the Second World War Konfin I mass grave in Slovenia. Int J Legal Med 124:307–317
29. Cattaneo C, Craig OE, James NT, Sokol RJ (1997) Comparison of three DNA extraction methods on bone and blood stains up to 43 years old and amplification of three different gene sequences. J Forensic Sci 42: 1126–1135
30. Bender K, Schneider PM, Rittner C (2000) Application of mtDNA sequence analysis in forensic casework for the identification of human remains. Forensic Sci Int 113:103–107
31. Hochmeister MN, Budowle B, Borer UV, Eggmann U, Comey CT, Dirnhofer R (1991) Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains. J Forensic Sci 36:1649–1661
32. Loreille OM, Diegoli TM, Irwin JA, Coble MD, Parsons TJ (2007) High efficiency DNA extraction from bone by total demineralization. Forensic Sci Int Genet 1:191–195
33. Jakubowska J, Maciejewska A, Pawlowski R (2012) Comparison of three methods of DNA extraction from human bones with different degrees of degradation. Int J Legal Med 126:173–178
34. Amory S, Huel R, Bilic A, Loreille O, Parsons TJ (2012) Automatable full demineralization DNA extraction procedure from degraded skeletal remains. Forensic Sci Int Genet 6:398–406
35. Zupanič-Pajnič I (2011) Highly efficient DNA extraction method from skeletal remains. Zdravniski Vestnik 80:171–181
36. Zupanc T, Balazic J, Stefanic B, Zupanic Pajnic I (2013) Performance of the human quantifiler, the investigator quantiplex and the investigator ESSplex plus kit for quantification and nuclear DNA typing of old skeletal remains. Rom J Leg Med 21:119–124
37. Zupanič-Pajnič I, Gornjak Pogorelc B, Balažic J, Zupanc T, Štefanič B. Highly efficient nuclear DNA typing of the World War II skeletal remains using three new autosomal short tandem repeat amplification kits with the extended European Standard Set of loci. Croat Med J. 2012; 53(1):17–23
38. Zupanič-Pajnič I (2013) A comparative analysis of the AmpF1STR identifier and PowerPlex 16 autosomal short tandem repeat (STR) amplification kits on the skeletal remains excavated from Second World War mass graves in Slovenia. Romanian Journal of Legal Medicine 21:73–78
39. Boom R, Sol CJA, Salimans MMM, Jansen CL, Wertheimvandillen PME, Vandernoordaa J (1990) Rapid and simple method for purification of nucleic-acids. J Clin Microbiol 28:495–503
40. Kishore R, Hardy WR, Anderson VJ, Sanchez NA, Buoncristiani MR (2006) Optimization of DNA extraction from low-yield and degraded samples using the BioRobot (R) EZ1 and BioRobot (R) M48. J Forensic Sci 51:1055–1061
41. Montpetit SA, Fitch IT, O'Donnell PT (2005) A simple automated instrument for DNA extraction in forensic casework. J Forensic Sci 50:555–563
42. Nagy M, Otremba P, Kruger C, Bergner-Greiner S, Anders P, Henske B et al (2005) Optimization and validation of a fully automated silica-coated magnetic beads purification technology in forensics. Forensic Sci Int 152:13–22
43. Valgren C, Wester S, Hansson O (2008) A comparison of three automated DNA purification methods in forensic casework. Forensic Sci Int Genet 1:76–77
44. Zupanič-Pajnič I, Debska M, Gornjak-Pogorelc B, Vodopivec Mohorčič K, Balažic J, Zupanc T et al (2016) Highly efficient automated extraction of DNA from old and contemporary skeletal remains. J Forensic Legal Med 37:78–86
45. Hofreiter M, Serre D, Poinar HN, Kuch M, Paabo S (2001) Ancient DNA. Nat Rev Genet 2:353–359
46. Handt O, Richards M, Trommsdorff M, Kilger C, Simanainen J, Georgiev O et al (1994) Molecular-genetic analyses of the Tyrolean Ice Man. Science 264:1775–1778
47. Kolman CJ, Tuross N (2000) Ancient DNA analysis of human populations. Am J Phys Anthropol 111:5–23
48. Handt O, Krings M, Ward RH, Paabo S (1996) The retrieval of ancient human DNA sequences. Am J Hum Genet 59:368–376
49. Wandeler P, Smith S, Morin PA, Pettifor RA, Funk SM (2003) Patterns of nuclear DNA

degeneration over time - a case study in historic teeth samples. Mol Ecol 12:1087–1093

50. Brown TA, Brown KA (1992) Ancient DNA and the archaeologist. Antiquity 66:10–23
51. Graham EAM (2007) DNA reviews: ancient DNA. Forensic Sci Med Pathol 3:221–225
52. Paabo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N et al (2004) Genetic analyses from ancient DNA. Annu Rev Genet 38:645–679
53. Pääbo S. Amplifying ancient DNA. In PCR-Protocols and Amplifications-A Laboratory Manual, ed. MA Innis, DH Gelfand, JJ Sninsky, TJ White, 1990, pp 159–66. San Diego: Academic
54. Wilson MR, Dizinno JA, Polanskey D, Replogle J, Budowle B (1995) Validation of mitochondrial-DNA sequencing for forensic casework analysis. Int J Legal Med 108:68–74
55. Bar W, Brinkmann B, Budowle B, Carracedo A, Gill P, Holland M et al (2000) DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. Int J Legal Med 113:193–196
56. Carracedo A, Bar W, Lincoln P, Mayr W, Morling N, Olaisen B et al (2000) DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int 110:79–85
57. Kalmar T, Bachrati CZ, Marcsik A, Rasko I (2000) A simple and efficient method for PCR amplifiable DNA extraction from ancient bones. Nucleic Acids Res 28:e67
58. Tully G, Bar W, Brinkmann B, Carracedo A, Gill P, Morling N et al (2001) Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. Forensic Sci Int 124:83–91
59. Alonso A, Andelinovic S, Martin P, Sutlovic D, Erceg I, Huffine E et al (2001) DNA typing from skeletal remains: evaluation of multiplex and megaplex STR systems on DNA isolated from bone and teeth samples. Croat Med J 42:260–266
60. Davoren J, Vanek D, Konjhodzic R, Crews J, Huffine E, Parsons TJ (2007) Highly effective DNA extraction method for nuclear short tandem repeat testing of skeletal remains from mass graves. Croat Med J 48:478–485
61. Tamariz J, Voynarovska K, Prinz M, Caragine T (2006) The application of ultraviolet irradiation to exogenous sources of DNA in plasticware and water for the amplification of low copy number DNA. J Forensic Sci 51:790–794
62. Shaw K, Sesardic I, Bristol N, Ames C, Dagnall K, Ellis C et al (2008) Comparison of the effects of sterilisation techniques on subsequent DNA profiling. Int J Legal Med 122:29–33
63. Milos A, Selmanovic A, Smajlovic L, Huel RLM, Katzmarzyk C, Rizvic A et al (2007) Success rates of nuclear short tandem repeat typing from different skeletal elements. Croat Med J 48:486–493
64. Misner LM, Halvorson AC, Dreier JL, Ubelaker DH, Foran DR (2009) The correlation between skeletal weathering and DNA quality and quantity. J Forensic Sci 54:822–828
65. Edson S, Ross JP, Coble MD, Parsons TJ, Barritt SM (2004) Naming the dead - confronting the realities of rapid identification of degraded skeletal remains. Forensic Sci Rev 16:64–89
66. Mundorff AZ, Bartelink EJ, Mar-Cash E (2009) DNA preservation in skeletal elements from the world trade center disaster: recommendations for mass fatality management. J Forensic Sci 54:739–745
67. Mundorff A, Davoren JM (2014) Examination of DNA yield rates for different skeletal elements at increasing post mortem intervals. Forensic Sci Int Genet 8:55–63
68. Schwartz TR, Schwartz EA, Mieszerski L, Mcnally L, Kobilinsky L (1991) Characterization of deoxyribonucleic-acid (DNA) obtained from teeth subjected to various environmental-conditions. J Forensic Sci 36:979–990
69. Sweet D, Hildebrand D (1998) Recovery of DNA from human teeth by cryogenic grinding. J Forensic Sci 43:1199–1202
70. Salamon M, Tuross N, Arensburg B, Weiner S (2005) Relatively well preserved DNA is present in the crystal aggregates of fossil bones. Proc Natl Acad Sci U S A 102:13783–13788
71. Sampietro ML, Gilbert MTP, Lao O, Caramelli D, Lari M, Bertranpetit J et al (2006) Tracking down human contamination in ancient human teeth. Mol Biol Evol 23:1801–1807
72. Gilbert MTP, Rudbeck L, Willerslev E, Hansen AJ, Smith C, Penkman KEH et al (2005) Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy. J Archaeol Sci 32:785–793
73. Lindahl T (1993) Instability and decay of the primary structure of DNA. Nature 362:709–715

Chapter 8

The Development and Use of Internal Amplification Controls (IACs) with DNA Profiling Kits for Forensic DNA Analysis

Nathalie Zahra and William Goodwin

Abstract

Biological samples recovered for forensic investigations are often degraded and/or have low amounts of DNA; in addition, in some instances the samples may be contaminated with chemicals that can act as PCR inhibitors. As a consequence this can make interpretation of the results challenging with the possibility of having partial profiles and false negative results. Because of the impact of DNA analysis on forensic investigations, it is important to monitor the process of DNA profiling, in particular the amplification reaction. In this chapter we describe a method for the in-house generation and use of internal amplification controls (IACs) with DNA profiling kits to monitor the success of the PCR proces. In the example we show the use of the SGM Plus® kit. These controls can also be used to aid the interpretation of the DNA profile.

Key words DNA profiling, AmpF*l*STR® SGM Plus®, Internal Amplification Controls, IACs, Quality Control, PCR

1 Introduction

Samples collected from crime or mass disaster scenes can range from body fluids (e.g., blood, saliva, and semen), to soft tissue (e.g., muscle) and hard tissue (e.g., bone and teeth). The state in which they are recovered and submitted for DNA analysis is often far from ideal. They can be found in limited quantities (<100 pg) such as with fingerprints [1] and hair [2], degraded after being subjected to environmental conditions such as UV light, humidity, and fluctuations in temperature or contaminated with external substances such as humic acid from soil or dyes from clothing. These factors can reduce the amount of DNA available for amplification and/or affect the amplification process itself. The net results can be partial or no DNA profiles and the interpretation to what caused this may be difficult, risking false negative results. Broemeling et al. [3] estimated that 10 % or more of crime scene

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_8,

samples fail to generate a profile due to insufficient DNA or co-extraction of PCR-inhibiting contamination.

An approach to monitor the PCR performance during DNA profiling with human identification kits would increase the confidence in the result and assist with its interpretation of the profile. The introduction of internal amplification controls (IACs) is a method used to directly monitor the efficiency of the enzymatic reaction. Also referred to as PCR internal controls, an IAC is defined as a non-target DNA fragment added to the sample and co-amplified with the target DNA in the same tube [4–8]. With an IAC, a reference signal is always produced even when no target DNA is present, indicating the success of the PCR. When the IAC signal is below optimal or absent, it indicates reduced efficiency or failure of the amplification reaction.

Amplification of an IAC within the IAC-PCR system can be done either competitively (using the same primers as the target DNA) or noncompetitively (using different set of primers). In competitive amplification the concentration, size, and secondary structure of the IAC are crucial to limit competition with the target DNA, while in the noncompetitive system the amplification of the IAC is restricted by the concentration of IAC primers. If independent primers are used, competition is limited to PCR resources such as dNTPs and DNA polymerase.

The IAC fragment itself can either be synthetically generated or through molecular biology techniques from either another sources of DNA or from a different region of the DNA under investigation. Most of the mechanisms used to generate the IAC make use of PCR, such as the composite primer technique or the overlapping extension technique [9]. In all cases the fragments are generated separately before being added to the PCR assay with the target DNA. Also, independent of the design and the generation method of the IAC, the IAC-PCR systems need to be empirically validated before being applied to casework samples.

In the field of genetics, the use of IACs is considered common practice, particularly in assays designed for the detection of microorganisms within complex biological samples such as food [10–12], clinical, and environmental samples [13–15]. It is done to establish the presence of inhibitors as part of a quality control system. With regard to forensic application, this role is taken by the real-time quantitative PCR assay such as Quantifiler® human DNA amplification kit and Plexor® qPCR system; both of which make use of a synthetic IAC, called an Internal PCR Control (IPC). However, it has been shown that real-time PCR shows different susceptibility to inhibition than conventional PCR, mostly related to the mechanistic difference between the two systems [2]. Any molecules present with the DNA that can interfere with the binding of the probe to the DNA, with the 5′–3′ exonuclease activity of the polymerase or with the fluorescence of the dye, may

potentially inhibit the real-time PCR [16, 17] and can give false quantification results [18]. Also, extremely high concentrations of human DNA (>10 ng/μl) can suppress the amplification of the IPC. An undetermined IPC signal can affect the proper interpretation of result as it can be due to PCR inhibitors or high DNA concentration [19].

The first attempt at using an IAC with conventional PCR during forensic DNA profiling was described by Kihlgren et al. [20]. In this study a modified D3S1358 product was used as IAC to monitor the presence of inhibitors in a system based on the amplification of the D3S1358 developed for the quantification of DNA. Another study includes that presented by Wurmb-Schwark [21], which describes the development of a new multiplex reaction to assess the degree of DNA degradation in bone DNA. It involves the amplification of five nuclear DNA loci (amelogenin, THO1, vWA, DYS390, and DYS391), two mitochondrial DNA hypervariable regions (280 and 439 bp), and a 271 bp fragment from *Hydra vulgaris* as an IAC for PCR inhibition. Although these studies are good examples for the application of the IAC in forensic genetics, in practical terms they require an additional assay to the DNA profiling kit, which consumes more of the DNA sample that in many cases is limited. Qiagen recently developed the Investigator® 24plex Quality Sensor kit which is able to amplify 22 human DNA STR loci in a single reaction (see Qiagen.com for details). Included in the reaction there are also two internal PCR controls which are amplified with the STRs and used to assess the quality of the PCR. However, laboratories may be tied into using specific kits and so it may be more beneficial to incorporate IACs into the kits that are currently being used.

Here, we present a procedure that can be used to generate and apply IACs directly with the DNA profiling kits that are currently being used in any given laboratory. As a proof of principle the procedure was developed using the SGM Plus® kit, which at the time was widely used in the UK. This kit is a short tandem repeat (STR) multiplex assay that amplifies 11 loci (D3S1358, vWA, D16S539, D2S1338, D8S1179, D21S11, D18S51, D19S433, TH01, FGA, and amelogenin) with fluorescently labeled primers. The fragments amplified are then resolved by capillary electrophoresis. Because the size range covered by the SGM Plus® amplicons is ~100–354 bp, two IACs were designed flanking this range with sizes of 90 and 410 bp. The plasmid pBR322 was used to generate the IACs to avoid sequence homology with human DNA. The two IACs fragments are initially generated through PCR, purified, and then added to the SGM Plus® reaction mix (Fig. 1). Within the reaction, the IAC_{90} and IAC_{410} are amplified noncompetitively with ROX (6-Carboxy-X-rhodamine) labeled primers. This will allow the detection of the IACs on the red lane together with the internal size standard.

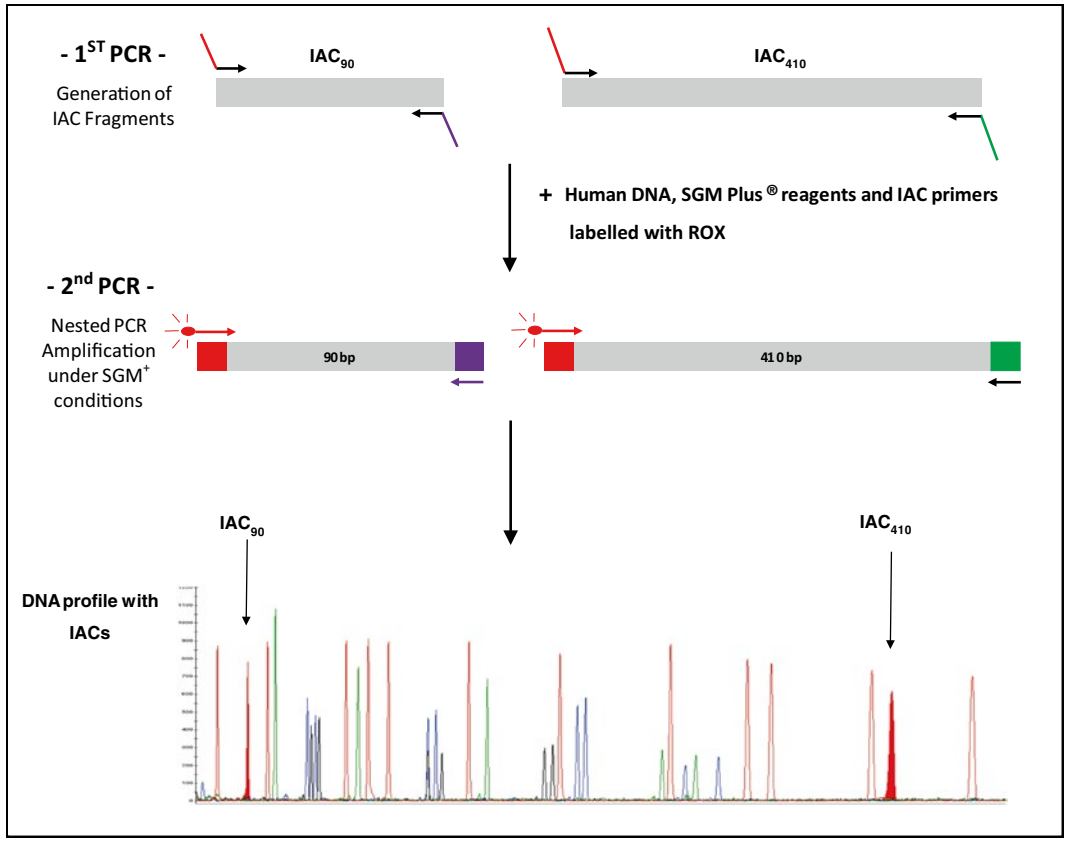


Fig. 1 Schematic diagram showing the process used for the generation and amplification of IACs. IAC_{90} and IAC_{410} fragments are generated using tailed primers (PCR 1). Fragments are then purified, added with the human DNA, and amplified noncompetitively using different primers complementary to the tail of the first set of primers (PCR 2). During PCR 2, the IAC fragments are amplified with a common ROX-labeled (➔) forward primer, allowing the detection of IACs markers with the human DNA profile

2 Materials

The following is a list of the reagents and instruments that are used for this protocol. Use good laboratory practice and follow laboratory guidelines when carrying out this procedure, particularly for the preparation of reagents, reactions, and waste disposal. Prepare all the solutions and reagents with ultrapure water (18 MΩ cm) and store at room temperature unless specifically indicated.

2.1 Reagents

- 0.5 M EDTA solution (pH 8.0) (Sigma Aldrich).
- 1.1× Ready Mix PCR Master Mix (ABgene).
- 1× TAE (prepared as described in Subheading 2.2).
- 6× loading buffer (ABgene).
- DNA 9947A (10 ng/μl) (Promega).

- Elution buffer (EB) (prepared as described in Subheading 2.2).
- Ethidium bromide (10 μg/ml) stock.
- Glacial acetic acid (Sigma Aldrich).
- GS™ 500 ROX™ Internal Standards (500 ROX™) (Applied Biosystems).
- Hi-Di formamide (Applied Biosystems).
- Isopropanol (Sigma Aldrich).
- Luria-Bertani (LB) media (*see* **Note 1**).
- POP-4™ polymer (Applied Biosystems).
- Penicillin (*see* **Note 1**).
- SeaKem® LE Agarose powder (Cambrex).

2.2 Preparation of Solutions

1. *50× TAE buffer*—Dissolve 242 g of Tris base in 750 ml 18 MΩ/cm water. Carefully add 57.1 ml glacial acetic acid and 100 ml 0.5 M EDTA (pH 8.0). Stir the mixture to dissolve the solids and add water for a final volume of 1 L. The solution can be stored at room temperature. One liter of 1× TAE buffer can be prepared by diluting 20 ml of 50× TAE in 980 ml of 18 MΩ/cm water.
2. *Elution buffer (EB)*—10 mM Tris–HCl (pH 8): For the preparation of 1 L, dissolve 121 g of Tris base in 900 ml of 18 MΩ/cm water. In a fume hood, adjust the pH to 8.0 with concentrated hydrochloric acid (HCl) using caution when handling the acid. Make the solution up to 1 L with 18 MΩ/cm water.

2.3 Kits Needed

The following kits are used during this procedure in our laboratory. Alternative kits can be used to perform the necessary task.

1. QIAquick® Gel extraction kit (Qiagen) containing:
   (a) QG buffer
   (b) PE Buffer
   (c) QIAquick column
2. Quant-iT™ PicoGreen® dsDNA kit (Life Technologies) containing:
   (a) Quant-iT™ PicoGreen® dsDNA reagent
   (b) 20× TE buffer
   (c) 100 μg/ml DNA standard
3. AmpF*l*STR® SGM Plus® kit (Applied Biosystems) containing:
   (a) AmpF*l*STR® PCR Reaction Mix
   (b) AmpF*l*STR® SGM Plus® Primer Mix
   (c) AmpF*l*STR® SGM Plus® Allelic Ladder
   (d) AmpF*l*STR® Control DNA 007
   (e) AmpliTaq Gold® DNA polymerase

2.4 Primers

Primers with sequences shown in Table 1 and Table 2 can be purchased from any molecular biological company. Before use:

1. Resuspend lyophilized primers with EB buffer to make a stock solution of 100 μM.
2. From this 100 μM stock solution prepare aliquots of 5 μM working solutions by diluting with EB buffer and use this working solution of the primers for the amplification reactions (*see* **Note 2**).

2.5 Plasmid pBR322 DNA

Plasmid pBR322 DNA can be ordered directly from molecular biology manufacturers. However a constant supply of the pBR322 DNA can be maintained through cloning of the vector in bacterium such as *Escherichia coli* and then performing the extraction of the pBR322 DNA when needed (*see* **Note 1**). Dilute down the plasmid DNA to 5 ng/μl.

2.6 Instruments Used

- PCR machine—A normal PCR instrument (of no particular manufacturer) is needed for the routine PCR reaction. However, for the optimization of the PCR conditions, a gradient PCR may be more appropriate.
- Electrophoresis chambers (various suppliers)
- UV Doc station
- TECAN Genio Pro™ plate reader
- ABI 310 Prism, ABI 3130 or 3500.

3 Methods

3.1 Generation of IACs

This part of the protocol gives practical details about the design of the IACs primers, the generation of the IAC_{90} and IAC_{410} template stock solutions, and the quantification and calculation of template copy number in the stock. The method is based on the use of

Table 1
The sequence of the primers used for the initial generation of the IAC_{90} and IAC_{410} template. The sequences in bold are the tails of the primers, which are used as binding sites for the primers amplifying the IACs during the SGM Plus® reaction

IAC fragment primer	Sequence
IAC_{90} Forward	**CTGTCAAATCTAAACACCCTGATGCG** GCTTGCGGTATTCGGAATCTTG
IAC_{90} Reverse	**GTCAGCTTGCATAATATCGAGATAACGC** GAGCGAGGGCGTGCAAGATT
IAC_{410} Forward	**CTGTCAAATCTAAACACCCTGATGCG** GATGCTGCTGGCTACCCTGT
IAC_{410} Reverse	**GTACAATGTTGACGTTCCTCGCTG** CGTGAAGCGATTCACAGATCTCTG

The sequences in bold are the tails of the primes, which are used as binding sites for the primers amplifying the IACs during the SGM Plus® reaction

Table 2
The primers sequences used for the amplification of the IAC_{90} and IAC_{410} during the SGM Plus® PCR. The forward primer is labelled with a ROX fluorescent dye at the 5' end. These primers bind to the tail of the IACs fragments previously generated

IAC fragment primer	Sequence
SGM Plus®IAC_{90} and IAC_{410} Forward	ROX - CTGTCAAATCTAAACACCCTGATGCG
SGM Plus®IAC_{90} Reverse	GTCAGCTTGCATAATATCGAGATAACGC
SGM Plus®IAC_{410} Reverse	GTACAATGTTGACGTTCCTCGCTG

pBR322 as the source of the DNA. Primers that we use are listed in Tables 1 and 2; however, it may be preferable to design your own primers.

3.1.1 Designing of IACs Primers

1. Select a region of interest from the pBR322 plasmid sequence and design primers for its amplification. This can be done either manually or using appropriate software such as Primer3Plus [22]. Ensure that the oligonucleotides generated follow the general criteria of primer design, i.e., 18–25 bp primer length, 40–60 % GC content, and 50–70 °C annealing temperature.
2. Ensure that the selected sequence of the plasmid and the primers that are going to be used for amplification show no complementarity to the human genome. This can be checked by inputting the sequences into the Basic Local Alignment Search Tool (BLAST) program and aligning them against the human genome. The sequences should show no significant similarity between the selected plasmid sequences and the human genome.
3. Add an 18–20 bp random sequence with 40–60 % GC content to the 5′ end of all the designed primers. These will act as the tail of the primers and be used as primer binding site for the second amplification with the DNA profiling kit. Make sure that the final size of the amplicon generated is the size you want to generate in the final amplification reaction (*see* **Note 3**).

3.1.2 Generation of the IACs

1. Prepare 5 μM working solution of the primers (Table 1) from the 100 μM stock solution using EB buffer (*see* **Note 4**).
2. To generate the IACs template, for each of the IAC prepare the following amplification reaction using 10 μl of 1 ng/μl of pBR322 DNA, 36 μl of 1.1× Ready Mix PCR Master Mix, and 2 μl of 5 μM forward and 2 μl of 5 μM reverse tailed primers. Add 18 MΩ/cm water to give a total volume of 50 μl (*see* **Note 5**).
3. Run the amplification reaction on a PCR block using the following conditions as guidelines: Initial incubation of 95 °C for 5 min; 30 cycles of 94 °C for 30 s, 58 °C for 1 min, and

72 °C for 1 min; final incubation of 72 °C for 45 min and hold at 15 °C.

4. Optimize the amplification reaction by systematically changing the annealing temperature and the concentration of the magnesium chloride ($MgCl_2$), primers concentration, and DNA quantity in the reaction. PCR products can be assessed by gel electrophoresis (*see* **Note 6**).
5. Use the optimize conditions to prepare multiple amplification reactions (five to ten 50 µl reactions) and generate the IACs fragments in bulk.
6. Purify and isolate the fragments using the procedure described in Subheading 3.1.3.

3.1.3 Purification of IAC Template

The purification step is required to remove any residual primers, nucleotides, DNA polymerase, salts, unspecific DNA products, and high molecular weight DNA that might interfere in downstream applications. A simple method for purification is by running the sample through a gel and isolate the desired amplicon from the gel after electrophoresis.

1. For each IAC fragment, combine the various PCR products (generated in Subheading 3.1.2) into one tube and add 0.3× the final PCR product volume of 6× loading buffer.
2. Prepare a 2 % agarose gel (*see* **Note 7**), load the entire pooled sample on multiple lanes, and run with 1× TAE buffer. Quickly visualize the gel on a UV documentation station (*see* **Note 8**).
3. Identify the amplicon of interest and using a sharp scalp cut the gel around the band. Care should be taken when handling the scalp and cutting should be done as close to the band as possible to avoid excess gel in the sample.
4. Extract the DNA from the gel using the QIAquick® Gel extraction kit. Combine the various sections of the gel containing the fragments into one or two 1.5 ml tubes, taking note of the total weight of the gel. For each 100 mg of gel, add 300 µl of buffer QG and incubate at room temperature until all the gel dissolves. Add one gel volume of isopropanol and vortex. Transfer the mixture to the QIAquick column and centrifuge for 1 min at 14,000 × *g*. Discard the flow through and wash the column twice with 500 µl of Buffer QG followed by 750 µl of Buffer PE each time discarding the flow through. Allow the column to dry for 2 min and then elute the DNA in two times 50 µl aliquots of EB buffer for a total 100 µl.
5. Quantify the stock IAC solution as described in Subheading 3.1.4, aliquot and store appropriately. Place at −20 °C for long-term storage (*see* **Note 9**).

3.1.4 Quantification of DNA

Various quantification methods can be used to determine the concentration of the DNA fragments. These range from UV, fluorescent spectrophotometer, or real-time PCR. The procedure described here makes use of the Quant-iT™ PicoGreen® dsDNA kit for the quantification of both the IACs and unknown DNA samples.

1. Start the procedure by diluting the 20× TE buffer down to 1× TE with 18 MΩ/cm water. Prepare also a working solution of the Quant-iT™ PicoGreen® dsDNA reagent by making a 200 times dilution with 1× TE buffer to get a DNA concentration of 2,000 ng/ml. Make enough volume to process all the samples, requiring 200 μl per sample.
2. For each Quant-iT™ PicoGreen® run, prepare two sets of standards—a low and high range DNA—as described in Tables 3 and 4, respectively, for a final volume of 400 μl (*see* **Note 10**).
3. For the IAC and unknown human DNA samples, use 3 μl of the sample with 197 μl of EB buffer. Then add 200 μl of PicoGreen® working solution for a final volume of 400 μl. To ensure reliability of the results, alongside the unknown samples, run DNA samples of known concentration, spiked samples (*see* **Note 11**), and blank (no DNA) samples.
4. Pipette 300 μl of the PicoGreen®-diluted DNA (standards, unknowns, and spiked samples) into the wells of a 96-well black plate. Place the standards into the first two rows of the plates with the highest concentration at position A1. Follow with the unknown samples and reference blanks.
5. Read the plate on an appropriate reader such as the TECAN Genio Pro™ plate reader (other plate readers can be used), measuring the fluorescent intensity at a wavelength of 520 nm.

Table 3
The volumes needed for the preparation of the low range DNA dilutions that are used for the Quant-iT™ PicoGreen run

Volumes (in μl)			Final DNA concentration (ng/ml)
DNA standard (2,000 ng/ml)	1× TE	PicoGreen®	
20	180	200	100
10	190	200	50
5	195	200	25
2	198	200	10
0.2	199.8	200	1
0	200	200	Blank

Table 4
The volumes needed for the preparation of the high-range DNA dilution that are used for the Quant-iT™ PicoGreen run

Volumes (in μl)			Final DNA concentration (ng/ml)
DNA standard (2,000 ng/ml)	1× TE	PicoGreen®	
200	0	200	1000
160	40	200	800
80	120	200	400
40	160	200	200
20	180	200	100
0	200	200	Blank

6. At the end of the run, use the raw readings of the DNA standard to construct the calibration curve, ensuring to have an R^2 value ≥0.99. Use this relationship to determine the concentration of the unknown samples in ng/ml and calculate the concentration of the original samples in ng/μl.

3.1.5 Calculation of Copy Number

1. To calculate the IAC copy number, use the following equation (for a worked example *see* **Note 12**):

$$C = \frac{m \times \mathrm{Ac}}{\mathrm{Mw} \times \mathrm{Gs}}$$

where:

C = copy number (c.n)
m = amount of DNA in grams
Ac = Avogadro's constant = 6.022×10^{23}/mol
Mw = mean weight of base pair = 649 Da
Gs = Genome size in bp

2. Prepare the IAC working solution with the appropriate copy number by diluting the solution with EB buffer.

3.2 Use of IACs with DNA Profiling Kit

3.2.1 Optimization of the IAC Amplification

Before applying the IACs to the SGM Plus® reaction (or other kits) with unknown samples, it is more appropriate to validate the use of the IACs with the reaction mix. This step can be carried out to determine and optimize the amplification efficiency of the IACs in the kit (e.g., SGM Plus®) reaction mix and to establish the optimal copy number (c.n.) of IAC to be added. Validation can initially be done using individual IAC and then the combination of the two.

1. Dilute the IACs with EB buffer down to 10,000, 1000, 100, and 10 c.n/μl by serial dilution.
2. For a 50 μl SGM Plus® reaction, add 21.0 μl of AmpF*l*STR® PCR Reaction Mix, 11 μl of AmpF*l*STR® PCR Primer Mix, 1 μl of AmpliTaq Gold DNA Polymerase, 1 μl of 5 μM ROX-labeled forward primer (Table 2), 1 μl of 5 μM reverse primer (Table 2) and 1 μl of IAC with the appropriate copy number (from step 1), 1 ng of DNA standard, and 18 MΩ/cm water for a final volume of 50 μl. Prepare the reactions in triplicate together with control samples containing human DNA only (without IACs) and IACs only (without human DNA) (*see* **Note 13**).
3. Perform the amplification reaction using the conditions provided by the manufacturer. For the SGM Plus® reaction use an initial incubation of 95 °C for 11 min; 28 cycles of 94 °C for 1 min, 59 °C for 1 min, and 72 °C for 1 min; final incubation of 60 °C for 45 min and hold at 10 °C.
4. After amplification, run the samples on a capillary electrophoresis (as described in Subheading 3.2.2). Assess the electropherograms for artifacts such as drop-ins, stutter, or heterozygous peak imbalance.
5. Determine the IAC copy number that gives an IAC signal between 500 and 2000 RFU (*see* **Note 14**) and empirically validate the system with decreasing amount of human DNA and a range of PCR inhibitors.
6. Repeat the procedure combining both of the IACs in a single reaction. Validate again for the effect of the IACs on the quality of the profiles generated and sensitivity of the reaction (as described in step 5).
7. Use the validated conditions with unknown human DNA samples recovered from scene of crime samples. Add the amount of DNA recommended by the manufacturer.

3.2.2 Running of the Samples with Capillary Electrophoresis

1. After amplification, mix 1 μl of the PCR reaction with 10 μl of Hi-Di formamide and 0.3 μl of 500 ROX™ internal size standard to give a total volume of 11.3 μl.
2. Together with each batch of samples, prepare also an allelic ladder by mixing 1 μl of the allelic ladder with 10 μl of Hi-Di formamide and 0.3 μl of 500 ROX™ standard.
3. Vortex the solutions and run the samples on a capillary electrophoresis platform such as ABI Prism® 310 and ABI Prism® 3130 and 3500 using POP-4™ polymer and Running Buffer (1×).

Table 5
Typical parameters for a capillary electrophoresis run with the SGM Plus™ reaction

Parameters	Value
Injection time	5 s
Injection voltage	15 kV
Run voltage	15 kV
Temperature	60 °C
Run time	30 min

4. Set up the electrophoresis run using the parameters shown in Table 5 as a guideline. The run time can be increased up to 35 min to ensure the detection of the IAC_{410}.
5. At the end of the run, analyze the electropherogram using the appropriate software such as GeneMapper® ID v3.2. Ensure that the right matrix file is used and specify ROX as the internal lane standard.

3.2.3 Interpretation of the Results Using IACs as Quality Control Markers

1. For each electropherogram check the quality of the internal lane standard in terms of peak heights, peak shape, and background noise level.
2. Starting with the "IACs only" control sample, identify the IAC_{90} and IAC_{410} peaks amongst the internal lane standard peaks and take note of their peak height. Use these values as references for the analysis of the other samples.
3. Analyze the rest of the unknown samples taking note first of the IACs markers followed by the rest of the peaks forming the human DNA profile. Interpret the profiles using the electropherograms in Fig. 2 as a guideline. Compare the peak heights of the IACs in the unknown samples with those of the control samples to monitor any significant difference.

4 Notes

1. If a stock culture of *Escherichia coli* containing the plasmid pBR322 is available, pipette 200 µl of culture into a 15 ml tube containing 2 ml Luria-Bertani (LB) medium with 100 µg/ml penicillin. Incubate the culture overnight at 37 °C with vigorous shaking. From the growth, transfer 1.5 ml of the culture into a 2 ml microcentrifuge tube and centrifuge at 12,000 × *g* for 30 s at 4 °C. Remove the medium by aspiration, leaving the pellet dry in the tube. Continue with the extraction of DNA from the plasmid as described by Sambrook et al. [23].

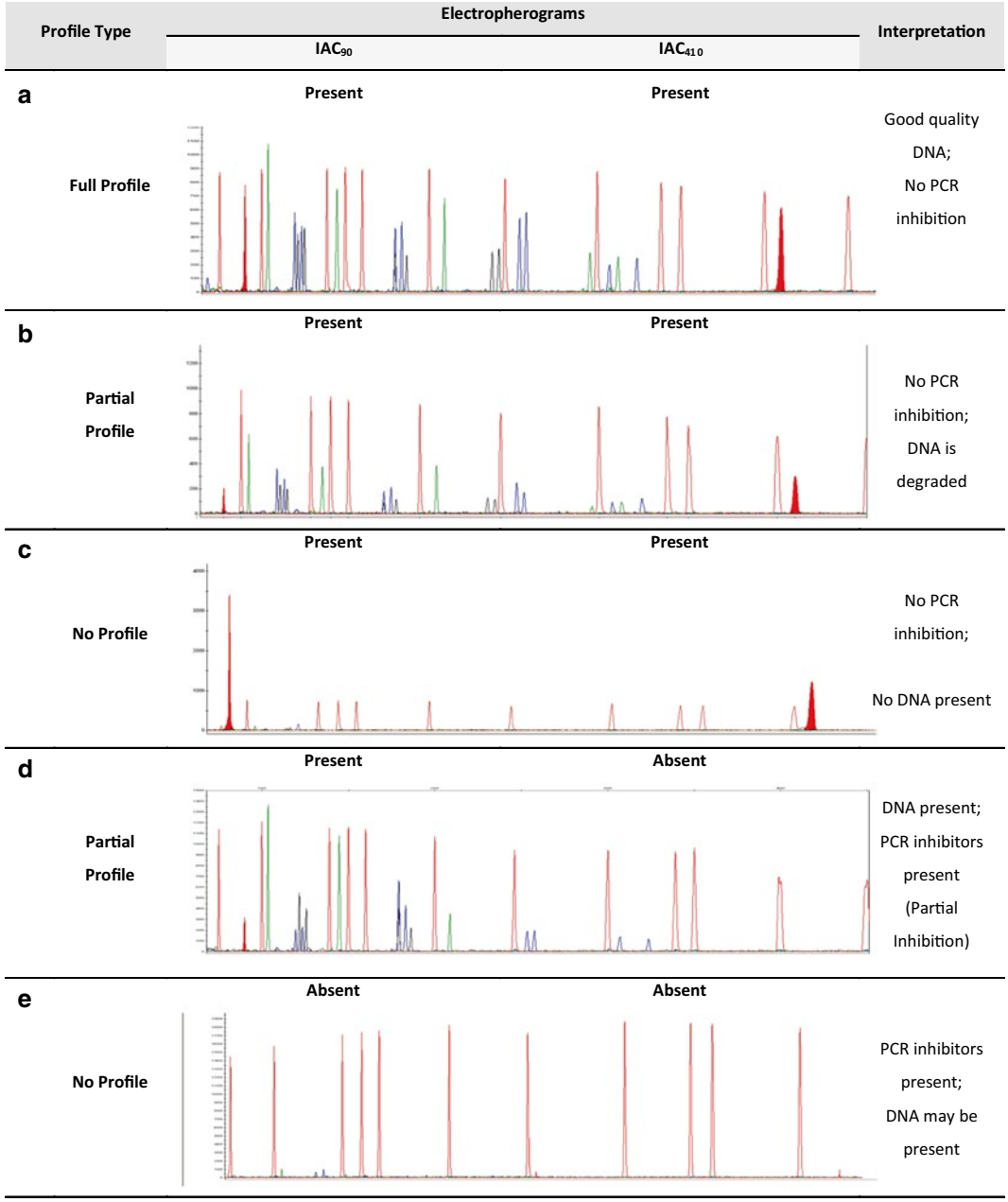


Fig. 2 Diagram showing the use of IACs markers during interpretation of DNA profiles. The system of markers is made up from IAC_{90} and IAC_{410}. When both markers are present together with the human DNA profile, it means there was no inhibition and injection was successful (**a**). However, when the signal of the human DNA is reduced or absent, the presence of the two markers suggests that human DNA is degraded (**b**) or absent (**c**), respectively. If the IAC markers are reduced along with the human DNA signal, it indicates the presence of inhibitors causing partial (**d**) or full inhibition (**e**). The lack of human DNA profile and IACs signal might imply that the sample was not injected

2. Primer stock solutions and working solutions can be stored at –20 °C. Fluorescently labeled primers should be kept away from direct light. Avoid freeze thawing of the solutions.
3. The size of the IACs generated after amplification with the DNA profiling kit should not overlap any of the size of the fragments forming the internal size standard used during capillary electrophoresis. This is to make sure both the IACs and internal size standard peaks are detected and identified during analysis.
4. Elution Buffer (EB) or water can be used to dilute DNA samples or primers or to prepare reagents. The use of TE (Tris-EDTA) buffer with 0.5–1 mM EDTA should be avoided as the EDTA can affect the amplification efficiency of the reaction. The EDTA can bind and sequester the magnesium ions that are needed by the polymerase to function.
5. PCR mixes such as the ABgene Master Mixes are ideal to use in this case. They include the appropriate concentration of dNTPs, polymerase, and magnesium chloride ($MgCl_2$) for standard PCR reactions. These master mixes only require the addition of the DNA template and primers. The reaction can however be set up using the individual components of the PCR.
6. Other methods, such as with a Bioanalyser, can be used for assessing the PCR products. It is important at this stage to make sure that the appropriate amplicon is being generated with no unspecific products.
7. 2 % agarose gel can be prepared by adding 2 g of agarose in 100 ml of 1× TAE buffer and then adding 3 μl of 10 μg/ml of ethidium bromide for a final concentration of 0.3 μg/ml. Care should be taken when handling ethidium bromide as it is known to be a mutagen. Volumes can be changed proportional to the volume of the gel needed while maintaining the ethidium bromide concentration in the range of 0.2–0.5 μg/ml.
8. Prevent long exposure of the DNA to UV as this can damage the DNA causing it to fragment and cross-link and making it unavailable for downstream processes.
9. Long-term storage of stock and diluted solution can be done in low binding DNA tubes and using EB buffer with 0.1 % Tween. This will prevent the DNA binding to the walls of the tube which can lead to a reduction in the concentration of the template with time.
10. When preparing DNA standards for a calibration curve, it is important to take extra care during pipetting. For a more accurate calibration curve, it is advisable to prepare the standards using the serial dilution method, each time vortexing and centrifuging the solution before preparing the next dilution.

11. A spiked sample can be prepared by adding a known amount of DNA (of known concentration) to an unknown DNA sample whose concentration is being measured in the same run as a separate (unspiked) sample. Calculating back the quantity of the spike from the raw fluorescent values will show the accuracy of the produce and give confidence in the rest of the readings generated during the run.
12. The following example shows the calculation of the DNA copy number in a solution. If we have an IAC_{90} stock solution with a concentration of 500 ng/μl (5×10^{-7} g/μl), the number of copies (C) in 1 μl is:

$$C = \frac{m \times \mathrm{Ac}}{\mathrm{Mw} \times \mathrm{Gs}}$$

$$C = \frac{\left(5 \times 10^{-7}\right) \times \left(6.022 \times 10^{23}\right)}{649 \times 90}$$

$$C = 5.15 \times 10^{13} \text{ copy numbers / μl}$$

13. When performing multiple samples, it is better to prepare a master mix containing all the components. In this case it will ensure that the IAC copy number is the same in all the samples. This will allow the samples to be directly compared with the controls during analysis and to determine the amplification efficiency.
14. Determining the optimal IACs copy number to use in the system is very important. Using a low copy number of the IACs can result in the random dropping out of the IACs signal, while using a disproportional high concentration of the IACs can result in the overamplification of the fragments. Too much IACs product can in turn negatively impact amplification reaction, reducing its efficiency in the amplification of the human DNA. In general the larger the size of the fragment, the higher the copy number that needs to be added.

References

1. van Oorschot RAH, Phelan DG, Furlong S, Scarfo GM, Holding NL, Cummins MJ (2003) Are you collecting all the available DNA from touched objects? Prog Forensic Genet 9 1239:803–807
2. Gilbert MTP, Menez L, Janaway RC, Tobin DJ, Cooper A, Wilson AS (2006) Resistance of degraded hair shafts to contaminant DNA. Forensic Sci Int 156:208–212
3. Broemeling DJ, Pel J, Gunn DC, Mai L, Thompson JD, Poon H et al (2008) An instrument for automated purification of nucleic acids from contaminated forensic samples. JALA 13:40–48
4. Cone RW, Hobson AC, Huang M-W (1992) Coamplified positive control detects inhibition of polymerase chain reactions. J Clin Microbiol 30:3185–3189
5. Ballagi-Pordány A, Belák S (1996) The use of mimics as internal standards to avoid false negatives in diagnostic PCR. Mol Cell Probes 10:159–164

6. Sachadyn P, Kur J (1998) The construction and use of a PCR internal control. Mol Cell Probes 12:259–262
7. Hoorfar J, Cook N, Malorny B, Wagner M, De Medici D, Abdulmawjood A et al (2003) Making internal amplification control mandatory for diagnostic PCR [2]. J Clin Microbiol 41:5835
8. Rodríguez-Lázaro D, D'Agostino M, Pla M, Cook N (2004) Construction strategy for an internal amplification control for real-time diagnostic assays using nucleic acid sequence-based amplification: development and clinical application. J Clin Microbiol 42:5832–5836
9. Hoorfar J, Malorny B, Abdulmawjood A, Cook N, Wagner M, Fach P (2004) Practical considerations in design of internal amplification controls for diagnostic PCR assays. J Clin Microbiol 42:1863–1868
10. Abdulmawjood A, Roth S, Bülte M (2002) Two methods for construction of internal amplification controls for the detection of escherichia coli O157 by polymerase chain reaction. Mol Cell Probes 16:335–339
11. Malorny B, Tassios PT, Rådström P, Cook N, Wagner M, Hoorfar J (2003) Standardization of diagnostic PCR for the detection of foodborne pathogens. Int J Food Microbiol 83:39–48
12. D'Agostino M, Wagner M, Vazquez-Boland JA, Kuchta T, Karpiskova R, Hoorfar J et al (2004) A validated PCR-based method to detect listeria monocytogenes using raw milk as a food model - towards an international standard. J Food Prot 67:1646–1655
13. Neumaier M, Braun A, Wagener C (1998) Fundamentals of quality assessment of molecular amplification methods in clinical diagnostics. Clin Chem 44:12–26
14. Stöcher M, Leb V, Berg J (2003) A convenient approach to the generation of multiple internal control DNA for a panel of real-time PCR assays. J Virol Methods 108:1–8
15. Ursi D, Dirven K, Loens K, Ieven M, Goossens H (2003) Detection of mycoplasma pneumoniae in respiratory samples by real-time PCR using an inhibition control. J Microbiol Methods 55:149–153
16. Abu Al-Soud W, Rådström P (2000) Effects of amplification facilitators on diagnostic PCR in the presence of blood, feces, and meat. J Clin Microbiol 38:4463–4470
17. Burkhart CA, Norris MD, Haber M (2002) A simple method for the isolation of genomic DNA from mouse tail free of real-time PCR inhibitors. J Biochem Biophys Methods 52:145–149
18. Ricci U, Marchi C, Previderè C, Fattorini P (2006) Quantification of human DNA by real-time PCR in forensic casework [cited 13 July 2015]
19. Koukoulas I, O'Toole FE, Stringer P, Van Oorschot RAH (2008) Quantifiler™ observations of relevance to forensic casework. J Forensic Sci 53:135–141
20. Kihlgren A, Beckman A, Holgersson S (1998) Using D3S1358 for quantification of DNA amenable to PCR and for genotype screening. Prog Forensic Genet 7 1167:31–33
21. von Wurmb-Schwark N, Preusse-Prange A, Heinrich A, Simeoni E, Bosch T, Schwark T (2009) A new multiplex-PCR comprising autosomal and y-specific STRs and mitochondrial DNA to analyze highly degraded material. Forensic Sci Int Genet 3:96–103
22. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M et al (2012) Primer3-new capabilities and interfaces. Nucleic Acids Res 40:e115
23. Sambrook J, Fritsch EF, Maniatis T (eds) (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Chapter 9

NucleoSpin® XS Columns for DNA Concentration and Clean-Up

William R. Hudlow

Abstract

The phenol–chloroform (organic) extraction method continues to be a preferred method for extraction of DNA from forensic evidence samples that may contain low quantities of DNA and polymerase chain reaction (PCR) inhibitors. The aqueous extracts from the organic extraction of DNA require subsequent concentration and cleanup, which has traditionally been performed with microdialysis filter units, including the Centricon® and Microcon® centrifugal filter devices. Here, we describe the use of the NucleoSpin® XS silica columns as an alternative for the concentration and purification of the aqueous extracts from the organic extraction and for the removal of PCR inhibitors from existing DNA extracts.

Key words DNA, DNA Extraction, PCR Inhibitor, Forensic

1 Introduction

The successful typing of DNA isolated from forensic evidence samples relies on the ability to generate extracts that are appropriately concentrated and relatively free of polymerase chain reaction (PCR) inhibitors. While semiautomated extraction methods [1–4] are currently available, the phenol–chloroform (organic)-based extraction method [5], with subsequent microdialysis filter concentration and cleanup of the DNA, is still employed in many laboratories. Concentration of the extracts from the phenol–chloroform extraction with microdialysis filtration units has the benefit of being a simple protocol that utilizes a single buffer; however, these devices may not remove common PCR inhibitors and the final volume of DNA extract can be highly variable. Thus, alternatives for concentrating and purifying aqueous extracts following phenol–chloroform extraction have been sought out by laboratories that continue to use the phenol–chloroform extraction. One such alternative is the silica-based MACHEREY-NAGEL NucleoSpin® XS column, which requires a short processing time (~30 min) and yields a consistently low final extract volume (~10 μL) that is

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_9,

compatible with most DNA typing methodologies. The NucleoSpin® XS columns can be used with aqueous extracts from standard organic extractions [6], existing extracts that are known to contain PCR inhibitors [7] and alkaline lysates [8].

The NucleoSpin® XS concentration and cleanup method is a simple, multistep process that includes binding the DNA to the silica membrane, washing the silica membrane to remove impurities, and eluting the DNA from the silica membrane. Specifically, the aqueous extract from the phenol–chloroform extraction is transferred to the NucleoSpin® XS column after the DNA binding conditions are adjusted with the addition of TE^{-4} buffer (10 mM Tris pH 7.5, 0.1 mM EDTA pH 8) and a guanidinium thiocyanate-based (NT) binding buffer. The silica membrane is washed with an ethanol-based wash buffer (B5) and the DNA is eluted in a low-salt buffer (BE). Finally, any residual ethanol from the wash step is removed from the extract by heating at 90 °C or 75 °C to yield either single-stranded or double-stranded DNA extracts, respectively.

2 Materials

1. NucleoSpin® gDNA cleanup XS kit (MACHEREY-NAGEL GmbH & Co KG).
   (NucleoSpin is a registered trademark of MACHEREY-NAGEL GmbH & Co KG).
2. Ethanol (95–100 %).
3. TE^{-4} buffer: 10 mM Tris pH 7.5, 0.1 mM EDTA pH 8.
4. 1.5 mL microcentrifuge tubes.
5. Prepare B5 wash buffer with the addition of 24 mL ethanol (95–100 %) prior to beginning the NucleoSpin® XS cleanup procedure (volumes may need to be adjusted depending on the size of the kit used).

Diligently follow all waste disposal regulations when disposing of waste materials.

3 Methods

3.1 NucleoSpin® XS Concentration and Cleanup Methods

3.1.1 NucleoSpin® XS Concentration and Cleanup of Extracts

The following method is based on a 420 µL organic extraction (e.g., 400 µL extraction buffer and 20 µL proteinase K). Volumes may be increased proportionally for larger volume extractions, but no more than 520 µL should be loaded onto the NucleoSpin® XS device at a time. All procedures are conducted at room temperature, unless otherwise noted.

1. Transfer the aqueous layer from the phenol–chloroform cleanup step of the organic extraction to a 1.5 mL tube containing 420 µL of TE^{-4} buffer and 210 µL of binding buffer (NT).

2. Vortex and spin briefly in a microcentrifuge.
3. Load approximately one-half of the DNA/TE^{-4}/NT solution into a NucleoSpin® XS filtration device in a 2 mL capless waste tube.
4. Spin for 30 s at ~11,000 × *g* in a microcentrifuge.
5. Discard waste in 2 mL tube and replace with a new 2 mL capless waste tube.
6. Load the remainder of the DNA/TE^{-4}/NT solution into the same NucleoSpin® XS filtration device, noting the orientation of the device in the centrifuge (e.g., hinge up).
7. Spin for 30 s at ~11,000 × *g* in a microcentrifuge. Additional loading steps may be conducted by repeating **steps 6** and **7**, if necessary as a result of an increased volume extraction (*see* **Note 1**).
8. Add 100 µL of wash buffer (B5) to the NucleoSpin® XS filtration device (*see* **Note 2**). The NucleoSpin® XS filtration device should be rotated ~180° from its orientation in the previous steps (e.g., hinge down versus hinge up).
9. Spin for 2 min at ~11,000 × *g* in a microcentrifuge.
10. Discard the waste in the 2 mL waste tube and replace with a labeled, 1.5 mL capless tube.
11. Add 6–15 µL (*see* **Note 3**) of elution buffer (BE) to the NucleoSpin® XS filtration device.
12. Spin for 1 min at ~11,000 × *g* in a microcentrifuge.
13. Add 6–15 µL (*see* **Note 3**) of elution buffer (BE) to the NucleoSpin® XS filtration device.
14. Spin for 1 min at ~11,000 × *g* in a microcentrifuge.
15. Remove the NucleoSpin® XS column from the 1.5 mL tube. Concentrate and remove residual ethanol from the DNA extract in the 1.5 mL tube by heating for:
    a. ~8 min in a 90 °C dry-heat bath (DNA will be single-stranded).

    or

    b. ~17 min in a 75 °C dry-heat bath (DNA will be double-stranded).
16. Transfer DNA extract to a storage tube, noting the approximate volume recovered.

3.1.2 NucleoSpin® XS Cleanup of Existing DNA Extracts

The following method is based on the cleanup of DNA extracts with volumes less than 800 µL. As DNA binding to the silica columns is pH dependent, with an optimal pH range of approximately 5–6, some extracts (e.g., Chelex) may require pH adjustment prior to

loading the NucleoSpin® XS columns. Volumes may be increased proportionally for larger volume extracts, but no more than 520 µL should be loaded onto the NucleoSpin® XS device at one time. Conduct all procedures at room temperature unless otherwise noted.

1. Bring the DNA extract to a final volume of 800 µL with the addition of TE^{-4} buffer.
2. Vortex and spin briefly in a microcentrifuge.
3. Transfer 200 µL of binding buffer (NT) to the diluted DNA sample.
4. Vortex and spin briefly in a microcentrifuge.
5. Load approximately one-half of the DNA/TE^{-4}/NT solution into a NucleoSpin® XS filtration device in a 2 mL capless waste tube.
6. Spin for 30 s at ~11,000 × *g* in a microcentrifuge.
7. Discard the waste in 2 mL tube and replace with a new 2 mL capless waste tube.
8. Load the remainder of the DNA/TE^{-4}/NT solution into the NucleoSpin® XS filtration device, noting the orientation of the device in the centrifuge (e.g., hinge up).
9. Spin for 30 s at ~11,000 × *g* in a microcentrifuge (*see* **Note 4**).
10. Add 100 µL of wash buffer (B5) to the NucleoSpin® XS filtration device (*see* **Note 2**). The NucleoSpin® XS filtration device should be rotated ~180° from its orientation in previous steps (e.g., hinge down versus hinge up).
11. Spin for 2 min at ~11,000 × *g* in a microcentrifuge.
12. Discard the waste in the 2 mL waste tube and replace with a labeled, 1.5 mL capless tube.
13. Add 6–15 µL (*see* **Note 3**) of elution buffer (BE) to the NucleoSpin® XS filtration device.
14. Spin for 1 min at ~11,000 × *g* in a microcentrifuge.
15. Add 6–15 µL (*see* **Note 3**) of elution buffer (BE) to the NucleoSpin® XS filtration device.
16. Spin for 1 min at ~11,000 × *g* in a microcentrifuge.
17. Remove the NucleoSpin® XS column from the 1.5 mL tube. Concentrate and remove residual ethanol from the DNA extract in the 1.5 mL tubes by heating for:
    (a) ~8 min in a 90 °C dry-heat bath (DNA will be single -stranded).

    or

    (b) ~17 min in a 75 °C dry-heat bath (DNA will be double-stranded).
18. Transfer DNA extract to a storage tube, noting the approximate volume recovered.

4 Notes

1. Additional loading steps may be conducted by repeating **steps 6** and 7, if necessary as a result of an increased volume extraction and the NucleoSpin devices have been successfully loaded 12 times with final extract volumes of approximately 6 mL.
2. To prevent possible carryover of residual guanidinium thiocyanate (a known PCR inhibitor) from the NT binding buffer that may have collected at the back of the NucleoSpin® XS column the NucleoSpin® XS filtration device should be rotated ~180° from its orientation in the previous steps (e.g., hinge down versus hinge up).
3. Typically, 10 μL elution volumes are sufficient for most typing methods, but the user may choose to use elution volumes as low as 6 μL to ensure the final extract volume is less than 10 μL after the final heating step or may choose to elute in volumes as high as 15 μL to ensure a sufficient volume of extract is available for multiple typing methods.
4. Additional loading steps may be conducted by repeating **steps 8** and **9** to accommodate large volume extracts and the NucleoSpin devices have been successfully loaded 12 times with final extract volumes of approximately 6 mL.

References

1. Greenspoon SA, Ban JD, Sykes K, Ballard EJ, Edler SS, Baisden M et al (2004) Application of the BioMek (R) 2000 laboratory automation workstation and the DNA IQ (TM) system to the extraction of forensic casework samples. J Forensic Sci 49:29–39
2. Cowan C, Schwandt M, Tambasco AJ (2009) Small scale automation in the forensic DNA laboratory. Forensic Mag 6:19–23
3. Brevnov M, Mundt J, Benfield J, Treat-Clemons L, Kalusche G, Meredith J et al (2009) Automated extraction of DNA from forensic sample types using the PrepFiler automated forensic DNA extraction kit. J Forensic Sci 14:294–302
4. Fregeau CJ, Lett CM, Fourney RM (2010) Validation of a DNA IQ (TM)-based extraction method for TECAN robotic liquid handling workstations for processing casework. Forensic Sci Int Genet 4:292–304
5. Sambrook J, Fritsch EF, Maniatis T (eds) (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
6. Hudlow WR, Krieger R, Meusel M, Sehhat JC, Timken MD, Buoncristiani MR (2011) The NucleoSpin(R) DNA clean-up XS kit for the concentration and purification of genomic DNA extracts: an alternative to microdialysis filtration. Forensic Sci Int Genet 5:226–230
7. Faber KL, Person EC, Hudlow WR (2013) PCR inhibitor removal using the NucleoSpin (R) DNA clean-up XS kit. Forensic Sci Int Genet 7:209–213
8. Hudlow WR, Buoncristiani MR (2012) Development of a rapid, 96-well alkaline based differential DNA extraction method for sexual assault evidence. Forensic Sci Int Genet 6:1–16

Chapter 10

Purification of PCR Products to Improve STR Profiles

Amy D. Roeder

Abstract

Forensic laboratories routinely perform STR analyses using commercially available STR kits. Very low levels of DNA are extracted from many forensic samples. In these samples, the amount of DNA that can be placed in a PCR is below the optimal DNA range for the commercial kits, leading to weak profiles and allelic dropout. STR profiles generated from poor quality samples can be improved by purification of the PCR product prior to capillary electrophoresis.

Key words Forensic science, DNA, STR profile, Post-PCR purification, Capillary electrophoresis

1 Introduction

Obtaining STR profiles from samples containing minimal quantities of DNA is a major challenge for forensic laboratories. A number of techniques, including increasing the number of PCR cycles, purification of the PCR amplicons prior to capillary electrophoresis, and increasing the time and voltage used for electro-injecting the samples during capillary electrophoresis have been described to improve the STR profiling success from low quality samples [1–9]. Increasing the number of PCR cycles is straightforward, but can make STR profile interpretation more complicated because of increased background noise and the presence of pull-up peaks resulting from the overamplification of some alleles. Increasing the electro-injection voltage and/or time is also straightforward, but can result in broadened peaks and increased background noise.

When a sample is electro-injected for capillary electrophoresis, salts, unincorporated dNTPs, primers, primer dimers, and PCR amplicons are transferred into the capillary. Decreasing the concentration of these unwanted components in the samples by post-PCR purification allows more PCR amplicons to be injected. In addition, the resultant electropherograms tend to be cleaner due to removal of dye-artifacts and labeled primers. When profiles from the same sample are compared pre- and post- purification, the relative

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_10,

fluorescent unit (rfu) values of the amplicon peaks are higher in electropherograms generated using purified PCR products.

Post-PCR purification is advantageous in that the sample can be amplified and electrophoresed under the laboratory's standard conditions. The remaining PCR products from samples that have weak profiles can be purified and electrophoresed under standard laboratory conditions or with increased electro-injection voltage and/or time. A post-PCR purification method and two methods to prepare the purified sample for capillary electrophoresis are described. The methodology described below is designed to increase the profiling success in AmpFlSTR® SGM Plus® and Identifiler® PCR Amplification Kits (Life Technologies) but can, in principle, be used for any PCR products.

2 Equipment and Materials

2.1 General Supplies and Equipment

- Amicon® Ultra 0.5 centrifugal filter unit with Ultracel-30 membrane (Merck Millipore).
- Collection tubes (provided with the filter units).
- 96-well reaction plate (suitable for capillary electrophoresis).
- 96-well plate septa (Life Technologies).
- Reagent grade water.
- Hi-Di™ formamide (Life Technologies).
- GeneScan® 500 Liz™ size standard (Life Technologies).
- GeneScan® 400 Rox™ size standard (Life Technologies).
- AmpFlSTR® Identifiler® PCR Amplification Kit (Life Technologies).
- AmpFlSTR® SGM Plus® PCR Amplification Kit (Life Technologies).
- Microcentrifuge.
- 96-well plate centrifuge.
- Vortex mixer.
- Thermal cycler.
- 3130 or 3500 Genetic Analyzer (Life Technologies).

3 Methods

3.1 Purification of PCR Products Using Amicon® Filter Units

The protocol is based on the instructions supplied by the manufacturer of the kit.

1. Place a filter unit, closed end downward, inside a collection tube.
2. The total volume of water + PCR products to be added to the filter unit should equal 500 μl. Calculate the amount of water required and add that amount to the filter unit (*see* **Note 1**).
3. Transfer PCR products to the filter unit.

4. Centrifuge the filter device/collection tube at 14,000 × *g* for 10 min.
5. Remove the filter device. Invert the filter device and place in a new collection tube.
6. Centrifuge the filter device/collection tube at 1000 × *g* for 2 min.
7. Discard the filter unit. The fluid in the collection tube (~20 µl) contains the purified PCR amplicons.

3.2 Capillary Electrophoresis of Purified PCR Amplicons

1. The purified PCR amplicons can be electrophoresed using either of the following methods.
   (a) Prepare the samples for electrophoresis by adding 1 µl of purified PCR product to 9 µl of Hi-Di™ formamide/size standard mixture in a 96-well plate. Add 0.015 µl of GeneScan® 400 Rox™ per sample to SGM Plus® PCR amplicons and 0.01 µl GeneScan® 500 Liz™ per sample to Identifiler® PCR amplicons. Add 0.054 µl of GeneScan® 400 Rox™ (SGM Plus®) or 0.2 µl of GeneScan® 500 Liz™ (Identifiler®) to the allelic ladder (not purified) (*see* **Note 2**).
   (b) Prepare the samples for electrophoresis by adding 9 µl of purified PCR product to 11 µl of Hi-Di™ formamide/size standard mixture in a 96-well plate. Add 0.03 µl of GeneScan® 400 Rox™ per sample to SGM Plus® PCR products and 0.01 µl GeneScan® 500 Liz™ per sample for Identifiler® PCR products. The allelic ladder is prepared for electrophoresis as in 1a.
2. Place the 96-well plate septa on the 96-well plate.
3. Centrifuge briefly to collect the liquid in the bottom of the wells and remove air bubbles.
4. Denature the DNA using a thermal cycler. Do not put the thermal cycler lid down on the septa.

   95 °C—3 min.
   4 °C—3 min.
   15 °C—∞.
5. Electrophorese the PCR products on a 3130 Genetic Analyzer using the following parameters (*see* **Note 3**):

   Oven Temperature: 60 °C.
   Pre-Run Voltage: 15 kV.
   Pre-Run Time: 180 s.
   Injection Voltage: 3 kV.
   Injection Time: 10 s.
   Voltage Number of Steps: 40.
   Voltage Step Interval: 15 s.
   Run Voltage: 15 kV.
   Run Time: 1350 s.

4 Notes

1. The water is used to decrease the salt concentration in the purified product.
2. An allelic ladder (provided with the STR kits) should be run with each electro-injection.
3. The method described has been validated for the 3130, if other platforms are used optimization will be necessary.

References

1. Balogh MK, Burger J, Bender M, Schneider PM, Alt KW (2003) STR genotyping and mtDNA sequencing of latent fingerprint on paper. Forensic Sci Int 137:188–195
2. Gill P (2001) Application of low copy number DNA profiling. Croat Med J 42:229–232
3. Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. Forensic Sci Int 112:17–40
4. Kloosterman AD, Kersbergen P (2003) Efficacy and limits of genotyping low copy number DNA samples by multiplex PCR of STR loci. Prog Forensic Genet 9:795–798
5. Lederer T, Braunschweiger G, Betz P, Seidl S (2002) Purification of STR-multiplex-amplified microsamples can enhance signal intensity in capillary electrophoresis. Int J Legal Med 116:165–169
6. Prinz M, Schiffner L, Sebestyen JA, Bajda E, Tamariz J, Shaler RC et al (2006) Maximization of STR DNA typing success for touched objects. Prog Forensic Genet 11:651–653
7. Roeder AD, Elsmore P, Greenhalgh M, McDonald A (2009) Maximizing DNA profiling success from sub-optimal quantities of DNA: a staged approach. Forensic Sci Int Genet 3:128–137
8. Smith PJ, Ballantyne J (2007) Simplified low-copy-number DNA analysis by post-PCR purification. J Forensic Sci 52:820–829
9. Whitaker JP, Cotton EA, Gill P (2001) A comparison of the characteristics of profiles produced with the AMPFlSTR(R) SGM plus(TM) multiplex system for both standard and low copy number (LCN) STR DNA analysis. Forensic Sci Int 123:215–223

Chapter 11

Analysis of 30 Biallelic INDEL Markers Using the Investigator DIPplex® Kit

Majid Bashir and Nur Haliza Bt Hassan

Abstract

Insertion/deletion polymorphisms (INDELs) are a relatively new class of a DNA marker to be used in forensic casework; used most commonly as a supplementary method to STR-based typing. INDELs, like SNPs, are particularly useful for the analysis of highly degraded DNA as the amplicon sizes are typically below 160 bp; they can also be valuable as an additional tool to help resolve kinship cases, with the advantage over STRs that they do not have high mutation rates. INDELs have an advantage over SNPs in that they are length polymorphisms and so can be analyzed by simply measuring the length of the allele(s). The Qiagen Investigator® DIPplex Kit is currently only one of two commercially available kits for the amplification of INDEL polymorphisms; it amplifies 30 biallelic INDEL loci and the amelogenin locus. The primers used are fluorescence labeled with 6-FAM, BTG, BTY, and BTR. This technique is robust, relatively simple, and the results are analyzed using the same capillary electrophoresis equipment and software as used for STR typing.

Key words INDEL, Investigator® DIPplex, Degraded DNA

1 Introduction

Currently, methods such as short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs) are used in forensic casework. STR analysis is the main tool utilized by the forensic community, but sometimes this technique gives unreliable results when profiling degraded samples as the amplicons sizes are relatively large at between 100 bp and 450 bp [1, 2]. To enhance the information recovered from degraded samples, miniSTRs have been developed with smaller amplicons; however, the amplicon sizes is still typically between 100 bp and 200 bp [3, 4] and profiling may not be successful when working with highly degraded DNA [5]. SNP typing is an alternative method for forensic identification from highly degraded DNA, with loci typically amplified on shorter amplicons than STR alleles. However, SNP analysis using profiling systems such as SNaPShot mini sequencing is a laborious and time-consuming method compared to STR analysis [6].

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_11,

With forensic analysis, INDELs share some advantages with both STRs and SNPs. The INDELs have short amplicons, typically 60 bp to 160 bp, and low mutation rates, making them suitable for the analysis of degraded DNA [7, 8] and as a useful supplementary tool for evaluating kinship tests [9]. Unlike SNPs, they are easy to analyze and require no additional steps post-PCR before analysis using capillary electrophoresis.

Thousands of insertion/deletion polymorphisms have been characterized in the human genome [10–12] and selected to produce multiplexes that have been developed for the analysis of INDELs. For example, a 38-plex was developed and applied successfully to the profiling of degraded samples where traditional STR were failed to generate complete DNA profiles [13]. The Investigator DIPplex® Kit was the first commercial kit for INDEL developed by Qiagen (Qiagen GmbH, Germany) [14]. This kit simultaneously amplifies 30 biallelic loci that are located over 19 autosomes (Table 1); all the INDELs selected in the kit are separated by at least 10 Mbp from commercially available STRs (Qiagen, Investigator DIPplex handbook). The evidential value of the 30 INDELs in comparison to other marker systems is shown in Table 2. INDEL multiplexes have also been developed for evaluating biological ancestry [15, 16].

Table 1
Locus-specific information of the Investigator® DIPplex Kit (adapted from Qiagen Handbook for Investigator® DIPplex Kit)

DIP locus	Chromosomal localization	GeneBank/ SNP ID	Motif (+DIP)	Reference allele
Blue	Xp22.1–22.3	M55418	TAAG	X
Amel. X	Yp11.2	M55419	CACG	Y
Amel. Y	7q31.1	rs1611048	TGGGCTTATT	+DIP
HLD77	2q31.1	rs2307959	AGCA	–DIP
HLD45	7q36.2	rs1611001	GCAGGACTGGGCACC	+DIP
HLD131	6q16.1	rs2307652	CACA	–DIP
HLD70	16q13	rs1610905	AGGA	–DIP
HLD6	17p11.2	rs1305047	TAAGT	–DIP
HLD111	5q14.1	rs1610937		+DIP
HLD58	4q25	rs2308292		+DIP
HLD56				
Green	20p11.1	rs16438	CCCCA	–DIP
HLD118	11q22.2	rs17174476	GTTT	–DIP
HLD92	12q22	rs2307570	ACTTT	–DIP
HLD93	14q23.1	rs2308163	TGAT	+DIP
HLD99	9q22.32	rs8190570	CCACAAAGA	–DIP
HLD88	15q26.1	rs2307433	GTAG	–DIP
HLD101	5q33.2	rs1305056	CTACTGAC	–DIP
HLD67				

(continued)

Table 1 (continued)

DIP locus	Chromosomal localization	GeneBank/ SNP ID	Motif (+DIP)	Reference allele
Yellow	8p22	rs2308072	AAGG	–DIP
HLD83	17p13.3	rs2307581	TCCTATTCTACTCTGAAT	–DIP
HLD114	2q11.2	rs28369942	GACTT	–DIP
HLD48	22q12.3	rs6481	GTGGA	–DIP
HLD124	21q22.11	rs8178524	GAAGTCTGAGG	–DIP
HLD122	22q11.23	rs16388	ATTGCC	–DIP
HLD125	5q12.3	rs1610935	GACAAA	+DIP
HLD64	7q21.3	rs17879936	GTAAGCATTGT	–DIP
HLD81				
Red	22q13.1	rs16363	TGTTT	–DIP
HLD136	3p22.1	rs2067235	CAACCTGGATT	–DIP
HLD133	13q12.3	rs17238892	AGAGAAAGCTGAAG	+DIP
HLD97	1p32.3	rs2307956	GGGACAGGTGGCCACTAGGAGA	–DIP
HLD40	1q31.3	rs2307924	ATTAAATA	–DIP
HLD128	1p22.1	rs17878444	CCTAAACAAAAATGGGAT	–DIP
HLD39	8q24.12	rs3081400	CTTTC	
HLD84				

HLD human locus DIP
–DIP deletion
+DIP insertion

Table 2
Comparison of CP values for different kits (adapted from Qiagen Handbook for Investigator® DIPplex Kit)

Kits	Loci	CPE/Trio[a]	CPI[b]	Population
Investigator DIPplex	30 DIPs	0.998	2.83×10^{-13}	German Caucasian
Investigator DIPplex	30 DIPs	0.9999	7.20×10^{-12}	Malaysian Malay
Investigator DIPplex	30 DIPs	0.9973	1.06×10^{-12}	Qatari
AmpFlSTR Minifiler	8 STRs	0.99998	8.21×10^{-11}	US Caucasian
AmpFlSTRSEfiler plus	11 STRs	1	7.46×10^{-14}	US Caucasian
Powerplex 16	15 STRs	1	5.46×10^{-18}	US Caucasian
Sanchez et al. [6]	52 SNPs	0.9998	5.00×10^{-21}	European Caucasian

[a]Combined probability of paternity exclusion
[b]Combined probability of identity

2 Materials

All components of the Investigator DIPplex Kit should be stored at –20 °C. Repeated thawing and freezing should be avoided and the Primer Mix and Allelic Ladder should not be exposed to light. As with all PCR-based systems, DNA samples and post-PCR

reagents (Allelic Ladder and DNA size standard) should be stored separately from the PCR reagents and the reaction mixtures should be prepared in an area separate from that used for DNA isolation and post-PCR analysis. The use of disposable barrier tips also helps to minimize cross-contamination.

2.1 Equipment and Materials

2.1.1 General Supplies and Equipment

1. Vortex mixer.
2. Microfuge.
3. Pipettes and barrier tips.
4. 0.2 ml thin-walled PCR tubes/strips.
5. Thermocycler.
6. Capillary electrophoresis platform (*see* **Note 1**).

2.2 Investigator® DIPPlex Kit Contents

The kit contains sufficient reagents for 100 reactions when using a 25 μl reaction volume. The components are:

Reaction Mix A (500 μl),
Primer Mix DIPplex (500 μl),
Multi Taq2 DNA polymerase (150 U),
Nuclease-free water (2 × 1.9 ml),
Control DNA 9948, 0.1 ng/μl (200 μl),
DNA size standard 550 (BTO) (50 μl), and
Allelic Ladder DIPplex.

3 Methods

As with all PCR-based procedures, great care must be taken to avoid contamination of the kit components and the samples. Separate pre- and post-PCR areas should be used in the different phases of the analysis. If analyzing low template amounts of DNA then extreme precautions have to be undertaken. We carry out all pre-PCR work in a laminar flow hood. In addition, all liquid transfers are carried out using barrier tips to minimize the potential of aerosol contamination.

3.1 PCR Amplification

1. Remove all components for the amplification from the freezer and allow to thaw.
2. Mix all tubes using a bench top vortex and pulse centrifuge.
3. Prepare a master mix as detailed in Table 3—prepare approximately 10 % more than is required (*see* **Note 2**).
4. Mix the master mix thoroughly and dispense into PCR tubes.
5. Add the appropriate amount of DNA and nuclease-free water to each reaction—the recommended quantity of DNA is

Table 3
Reaction setup (*see* Note 2)

Component	Volume per reaction		
	Full	One Fourth	One Fifth
Reaction Mix A*	5.0 µl	1.25 µl	1.0 µl
Primer Mix	5.0 µl	1.25 µl	1.0 µl
Multi Taq2 DNA Polymerase	0.6 µl	0.15 µl	0.12 µl
Nuclease-free water	9.4 µl	2.35 µl	1.88 µl
Template DNA	5.0 µl	1.25 µl	1.0 µl
Total volume	*25 µl*	*6.25 µl*	*5.0 µl*

* This contains dNTPs, MgCl and bovine serum albumin

0.2–0.5 ng, but good quality results can be obtained with sub-optimal amounts (*see* **Note 3**).

6. Prepare positive (9948 or any other available sample of known concentration) and negative controls (nuclease-free water) (*see* **Note 4**).
7. Vortex and centrifuge briefly the samples to collect contents at the bottom of the tubes before adding it to the master mix.
8. Program the thermal cycler according to the manufacturer's instructions, using the conditions outlined in Table 4 (*see* **Note 5**).
9. Once the protocol is completed either store the samples at –20 °C or proceed directly with capillary electrophoresis.

3.2 Capillary Electrophoresis

Analysis of the PCR products should take place in an area that is physically separated from the pre-PCR area.

The analysis of the amplicons is essentially the same as with all other fragment analysis. A matrix file will have to be created using the BT5 dyes with filter set G5: 6-FAM, BTG, BTY, BTR, and BTO.

1. Prepare a mixture with Hi-Di formamide and GeneScan™-500 LIZ® size standard using 10 µl of Hi-Di Formamide and 0.5 µl of DNA Size standard 550 (BTO) or GeneScan™-500 LIZ® per sample (*see* **Note 6**).
2. Aliquot 10.5 µl of the mixture into each well.
3. Add 1 µl of PCR product or allelic ladder and mix thoroughly.
4. Denature the samples for 3 min at 95 °C and then snap freeze on ice for 3 min.
5. Load the samples onto the Genetic Analyzer. For the 3500 and 3130XL, use the following parameters: run temperature 60 °C,

Table 4
Standard PCR protocol for DIPplex Kit, Qiagen (adapted from Qiagen Handbook for DIPplex Kit)

Temperature	Time	Number of cycles
94 °C	4 min	1
94 °C 61 °C 72 °C	30 s 120 s 75 s	30
68 °C	60 min	1
10 °C	∞	

pre-run voltage 15 kV, pre-run time 180 s, injection time 10 s, injection voltage 1.6 kV, run voltage 15 kV, and run time 900 s. When using the ABI 310, use the following conditions: run temperature 60 °C, pre-run voltage 15 KV, pre-run time 180 s, injection time 5 s, injection voltage 1.6 kV, run voltage 15 kV, and run time 1440 s (*see* **Note 1**).

6. The data analysis can be done using Genemapper ID-X software using the settings recommended by manufacturer. Qiagen provide predefined bins on their Investigator® DIPplex template files (*see* **Note 7**). The exact length of the amplified products depends on the device type, the conditions of electrophoresis, as well as the DNA size standard used.
7. DIPsorter software can be used to create the INDEL profiles in PDF or CSV format (*see* **Note 8**). This involves conversion of Genemapper project into notepad (txt) and then DIPsorter software use this txt format to change into either PDF or CSV files which then can easily be used for interpretation and statistical calculations.

4 Notes

1. We have used the ABI310 with POP4, the 3130XL and 3500 (both with POP6) and have obtained consistent results. The run time with POP6 is 2500 s.
2. Based on our experience 1/5 volume yields good quality profiles without any major impact on peak balance, especially when profiling reference samples. Even when profiling casework samples reduced volumes produce stable profiles, as long as sufficient template DNA can be added in 1 μl of sample DNA (Fig. 1).
3. We have found that full profiles can be routinely obtained using as little as 100 pg of input DNA.

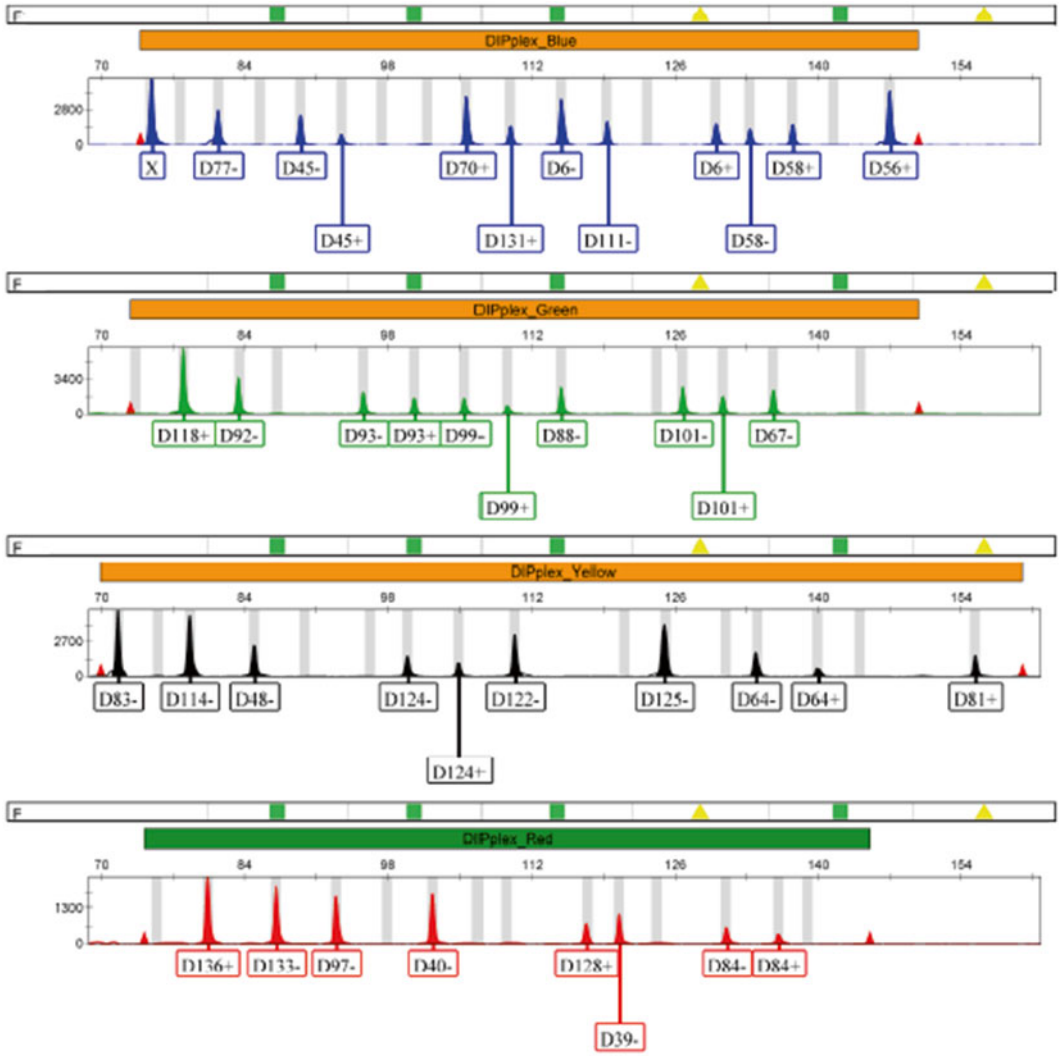


Fig. 1 Electropherogram of the Investigator DIPplex Kit using 200 pg of DNA recovered from the crime scene in Malaysia; the same extract could only be profiled at three loci when using the PowerPlex16 STR kit (Promega). Analysis was performed on an Applied Biosystems 3500 Genetic Analyzer. Allele assignment was performed using GeneMapper *ID-X* Software

4. The positive control provided with the kit (9948) will produce a full profile when only 1 μl of sample is added (100 pg).
5. The Handbook recommends using 32 cycles when working with low template DNA. However, we have found that using 32 cycles produced unacceptable levels of artifacts and so have always used 30 cycles. It should be noted that the samples that we have amplified using the Investigator® DIPplex Kit were not prepared under low template conditions.

6. In our hands when using the BTO 550 size standard provided with the kit, using either a 3130XL or a 3500, many of the alleles were called off-ladder. Therefore we adopted the use of GeneScan™-500 LIZ™. The filter set remains the same (i.e., G5) and the BT5 matrix set is also used.
7. The bin settings needed to be modify with reference to the allelic ladder to compensate for variation, the electrophoretic mobility of the fragments on different platforms and under different conditions (such as environmental temperature).
8. DIPsorter freeware available from Qiagen (https://www.qiagen.com).

References

1. Cotton EA, Allsop RF, Guest JL, Frazier RRE, Koumi P, Callow IP et al (2000) Validation of the AMPFlSTR® SGM plus(TM) system for use in forensic casework. Forensic Sci Int 112:151–161
2. Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ et al (2002) Validation of a 16-locus fluorescent multiplex system. J Forensic Sci 47:773–785
3. Butler JM, Shen Y, McCord BR (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. J Forensic Sci 48:1054–1064
4. Chung DT, Drábek J, Opel KL, Butler JM, McCord BR (2004) A study on the effects of degradation and template concentration on the amplification efficiency of the STR miniplex primer sets. J Forensic Sci 49:733–740
5. Zhang Z, Wang B, Guan H, Pang H, Xuan J (2009) A LDR-PCR approach for multiplex polymorphisms genotyping of severely degraded DNA with fragment sizes <100 bp. J Forensic Sci 54:1304–1309
6. Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M et al (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27:1713–1724
7. Fondevila M, Pereira R, Gusmão L, Phillips C, Lareu MV, Carracedo A et al (2011) Forensic performance of insertion-deletion marker systems. Forensic Sci Int Genet Suppl Ser 3:e443–e444
8. Manta F, Caiafa A, Pereira R, Silva D, Amorim A, Carvalho EF et al (2012) Indel markers: Genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material. Forensic Sci Int Genet 6:658–661
9. Pinto N, Magalhães M, Conde-Sousa E, Gomes C, Pereira R, Alves C et al (2013) Assessing paternities with inconclusive STR results: the suitability of bi-allelic markers. Forensic Sci Int Genet 7:16–21
10. Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G (2002) Human diallelic insertion/deletion polymorphisms. Am J Hum Genet 71:854–862
11. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS et al (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res 16:1182–1190
12. Montgomery SB, Goode DL, Kvikstad E, Albers CA, Zhang ZD, Mu XJ et al (2013) The origin, evolution, and functional impact of short insertion-deletion variants identified in 179 human genomes. Genome Res 23:749–761
13. Pereira R, Phillips C, Alves C, Amorim A, Carracedo Á, Gusmao L (2009) A new multiplex for human identification using insertion/deletion polymorphisms. Electrophoresis 30:3682–3690
14. LaRue BL, Ge J, King JL, Budowle B (2012) A validation study of the Qiagen Investigator DIPplex® kit; an INDEL-based assay for human identification. Int J Legal Med 126:533–540
15. Santos NPC, Ribeiro-Rodrigues EM, Ribeiro-dos-Santos ÂKC, Pereira R, Gusmão L, Amorim A et al (2010) Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. Hum Mutat 31:184–190
16. Zaumsegel D, Rothschild MA, Schneider PM (2013) A 21 marker insertion deletion polymorphism panel to study biogeographic ancestry. Forensic Sci Int Genet 7:305–312

Chapter 12

Analysis of Mitochondrial Control Region Using Sanger Sequencing

David Ballard

Abstract

The analysis of mitochondrial DNA (mtDNA) is an established forensic tool and has been used extensively to aid with both the identification of human remains and evidence recovered from scenes of crime. The biology of mtDNA confers both advantages and disadvantages when using it as a tool for identification. It benefits from a high copy number, which facilitates analysis from samples with highly degraded DNA or trace amounts of DNA, but the maternal mode of inheritance restricts its power of discrimination. With Next Generation Sequencing being used in research and some forensic casework laboratories the scope of mtDNA analysis in forensic casework may expand in the near future. Currently, however, most casework laboratories rely on Sanger sequencing and an established method for analyzing the hypervariable sequence regions is described.

Key words Mitochondrial DNA, Hypervariable sequence region I (HVS-I), Hypervariable sequence region II (HVS-II), Hypervariable sequence region III (HVS-III)

1 Introduction

Mitochondria are distinct organelles within the cell cytoplasm that play a crucial role in cellular respiration. Mitochondria have their own DNA content independent of the cell's nuclear DNA, comprising of approximately 16,569 base pairs arranged in a circular double stranded configuration [1]. The first mitochondrial sequence was published in 1981 by Anderson et al. [1], and can be found referred to as either the Anderson sequence or Cambridge Reference Sequence (CRS) while a subsequent reanalysis in 1999 resulted in 11 minor changes to this reference sequence. All human mitochondrial genomes are now compared with this Revised Cambridge Reference Sequence (rCRS) [2].

Unlike nuclear DNA, which is generally present in cells in just two copies (one copy of each chromosome from each parent), the copy number of mitochondrial DNA is much higher. In 1991, Satoh and Kuroiwa [3] experimentally determined that each

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_12,

mitochondrion within a cell contained 1–15 copies of the mitochondrial genome, at an average of 4–5 copies. Additionally each cell contains multiple mitochondria, giving an average copy number of 200–1700 [4] mitochondrial genomes per cell, although this is known to be tissue specific, varying between 50 and 75 copies in a spermatozoon [5] up to the order of 100,000 in a maturing primary oocyte [6]. The presence of multiple copies of the mitochondrial genome helps to explain why it can often be analyzed in situations where nuclear DNA has degraded (e.g., ancient bones). Furthermore, the cellular location of the mitochondrial DNA (in the mitochondria organelles) also has a strikingly protective function in regard to DNA degradation [7].

In a forensic scenario, mitochondrial analysis is most commonly performed in cases where nuclear DNA is unlikely to be present, notably in analysis of shed hairs and bones. Given the strictly matrilineal inheritance of the mitochondrial genome [5, 8], it can also be used to help resolve complex relationship cases where there is a common maternal link.

The more bases within the mitochondrial genome that are analyzed, the more discriminating the results will be. The advent of newer, so called, next generation sequencing technologies (e.g., Illumina's MiSeq or Life Technologies' Ion PGM) means that it is now relatively easy to sequence the entire mitochondrial genome of an individual, for example, using the PCR strategy described by Lyons et al. [9] (also *see* Chapter 12). Amplification of such long PCR amplicons relies on having reasonable quality DNA; however, this is still a feasible approach for some case samples, in addition to being a valid strategy when analyzing reference samples or for use in complex relationship cases.

A disproportionate amount of variation is centered within the 1.1 kb control region of the mitochondrial DNA and is even more concentrated around three areas within this control region termed hypervariable region I (HVS-I), hypervariable region II (HVS-II), and hypervariable region III (HVS-III). This is the area most traditionally sequenced for forensic analysis because a reasonable amount of information can be obtained while only sequencing a smaller fraction of the mitochondrial genome, saving time and money when using a traditional Sanger sequencing method, and providing a more relevant amplification target for degraded DNA samples.

While it is possible to amplify HVS-I alone, or additionally HVS-II and HVS-III, it is of most benefit to analyze the entire control region due to the increased information content contained within the whole genome. The amplification strategy for this will depend on the quality of the mitochondrial DNA extracted. Listed below are two protocols for amplifying the control region either in a single 1.1 kb amplicon or in 5 ~ 300 bp overlapping amplicons. Should the DNA be too degraded for even this approach to produce results, then it is possible to amplify the region in ten

amplicons ranging in size from 101 to 190 bp, details of which can be found in Eichmann et al. [10].

2 Materials

2.1 PCR

1. Extracted DNA (*see* **Note 1**).
2. MyTaq HS Red Mix 2× (Bioline B10-25047) (*see* **Note 2**).
3. Primers
   (a) Full control region.
      - L15977 [11] – cac cat tag cac cca aag ct
      - H599 [12] – ttg agg agg taa gct aca ta
   (b) Control region in ~300 bp amplicons [10].
      - L15989 – ccc aaa gct aag att cta at
      - H16248 – gtt gca gtt gat gtg tga tag
      - L16197 – ctt aca agc aag tac agc aat caa c
      - H16509 – agg aac cag atg tcg gat ac
      - L16450 – gct ccg ggc cca taa cac ttg
      - H180 – aat att gaa cgt agg tgc gat aaa t
      - L109 – gca ccc tat gtc gca gta tct gtc
      - H389 – gca ccc tat gtc gca gta tct gtc
      - L317 – ctt ctg gcc aca gca ctt aaa c
      - H619 – aac att ttc agt gta ttg ctt tga g
4. DNA-free water.
5. 2800M positive control DNA (Promega DD7101) (*see* **Note 3**).

2.2 Gel Electrophoresis

1. Analytical grade agarose (e.g., Promega V3125).
2. 10× TBE (e.g., Sigma T4415-1 L).
3. Gel Red nucleic acid stain (VWR 730–2958) (*see* **Note 4**).
4. DNA size ladder (e.g., HyperLadder 100 bp, Bioline BIO-33056).

2.3 Sequencing

1. BigDye terminator cycle sequencing v3.1 ready reaction kit (Life Technologies).
2. Sequencing primers. For the full control region PCR strategy, the following additional primers are required [12]:
   (a) H16401 - tga ttt cac gga gga tgg tg
   (b) L16450 - gct ccg ggc cca taa cac ttg
   (c) H274 - tgt gtg gaa agc ggc tgt gc

(d) L314 - ccg ctt ctg gcc aca gca ct

(e) H16175 - tgg att ggg ttt tta tgt a

3. Illustra ExoStar 1-Step (VWR US77705).
4. 125 mM EDTA—made up from EDTA disodium salt dihydrate (MW 372.24) powder (Sigma-Aldrich E4884). Dissolve 93 mg in 2 ml H_2O.
5. 100 % ethanol.
6. 70 % ethanol—made by combining absolute ethanol (100 %) with water in a ratio of 7:3.
7. Deionized formamide (Life Technologies).

3 Methods

3.1 Initial PCR

An initial PCR must be performed to amplify the region of interest: one of the two following PCR strategies should be followed.

If DNA is believed to be of reasonable quality then the full control region strategy (*see* Subheading 3.1.1) should be followed—this amplifies the entire control region in a single reaction with a 1.1 kb amplicon. If DNA is believed to be of poor quality, or the full control region strategy fails, then the Midi-Mito strategy (*see* Subheading 3.1.2) should be followed—this requires the use of two multiplex PCRs that amplify the control region in five smaller, overlapping, fragments.

It is possible to specifically quantify the amount of mitochondrial DNA present within the sample to analyze [11]; however, in the absence of a commercial solution, implementation of this method can be challenging. While specific mitochondrial DNA quantification is desirable, a viable alternative is to proceed to PCR without quantification, adding variable amounts of DNA depending on the sample type and expected cleanliness of the extract. It is assumed here that mitochondrial quantification has not been performed.

3.1.1 Full Control Region Strategy

1. PCR should be setup using DNA-free consumable in a pre-PCR laboratory environment. Reactions can be carried out in 0.2 ml PCR tubes or 96-well PCR plates as desired.
2. The required components and quantities for setting up the PCR are listed in Table 1. Reactions are performed in a 20 μl final volume.
3. A master mix should be prepared containing all the components except the DNA and water; 11.6 μl of this master mix can then be aliquoted into each PCR tube or well.

Table 1

Polymerase chain reaction reagent concentrations and volumes for amplification of the control region in one reaction

Reagent	Amount added per reaction (μL)
MyTaq HS Red Mix 2×	10.0
L15977 primer (5 μM)	0.8
H599 primer (5 μM)	0.8
DNA extract and water	8.4
Total	20.0

4. Up to 8.4 μl of extracted DNA can then be added to the reaction. For good quality samples or those likely to contain inhibitors, adding 2 μl of DNA and 6.4 μl of water is recommended.
5. A positive and negative control should be included with each batch: for the positive control add 1 ng of 2800 M DNA (*see* **Note 5**) while for the negative control substitute sample DNA with DNA-free water.
6. PCR should be carried out in a PCR machine with a heated lid using the following program: an initial activation of 95 °C for 3 min followed by 36 cycles of 94 °C for 15 s, 56 °C for 1 min, and 72 °C for 2 min.

3.1.2 Midi-Mito Strategy

1. Two multiplex PCR reactions must be setup for each sample. PCR should be carried out using DNA-free consumable in a pre-PCR laboratory environment. Reactions can be prepared in 0.2 ml PCR tubes or 96-well PCR plates as desired.
2. The required components and quantities for setting up Midi-Mito PCR A are listed in Table 2. Reactions are performed in a 20 μl final volume.
3. A master mix should be prepared containing all the components for Midi-Mito PCR A except the DNA and water; the appropriate volume (14.8 μl) of this master mix can then be aliquoted into each PCR tube or well.
4. Up to 5.2 μl of extracted DNA can be added to the reaction. For good quality samples or those likely to contain inhibitors, adding 2 μl of DNA and 3.2 μl of water is recommended.
5. For each sample, repeat **steps 2–4** above to setup Midi-Mito PCR B by using the PCR recipe detailed in Table 3 that produces a master mix volume of 13.2 μl and allows a maximum of 6.8 μl of DNA to be added for each sample.

Table 2

Polymerase chain reaction reagent concentrations and volumes for Midi-Mito PCR A amplifying three of the five Midi-Mito reactions

Reagent	Amount added per reaction (μL)
MyTaq HS Red Mix 2×	10.0
L15989 primer (5 μM)	0.8
H16248 primer (5 μM)	0.8
L16450 primer (5 μM)	0.8
H180 primer (5 μM)	0.8
L317 primer (5 μM)	0.8
H619 primer (5 μM)	0.8
DNA extract and water	5.2
Total	20.0

Table 3

Polymerase chain reaction reagent concentrations and volumes for Midi-Mito PCR B amplifying two of the five Midi-Mito reactions

Reagent	Amount added per reaction (μL)
MyTaq HS Red Mix 2×	10.0
L16197 primer (5 μM)	0.8
H16509 primer (5 μM)	0.8
L109 primer (5 μM)	0.8
H389 primer (5 μM)	0.8
DNA extract and water	6.8
Total	20.0

6. A positive and negative control should be included with each batch for both PCR A and PCR B: for the positive control add 1 ng of 2800 M DNA while for the negative control substitute sample DNA with DNA-free water.
7. PCR should be carried out in a PCR machine with a heated lid using the following program: an initial activation of 95 °C for 3 min followed by 34 cycles of 94 °C for 15 s, 57 °C for 10 s, and 72 °C for 20 s.

3.2 PCR Quality Check

Before proceeding to the sequencing stage, PCR products are first visualized on an agarose gel to check PCR success (*see* **Note 6**).

1. Make a 0.5× TBE working solution by diluting the 10× TBE stock 1 in 20 with water.
2. Prepare a 1.5 % agarose gel by dissolving analytical grade agarose into 0.5× TBE, e.g., to prepare 100 ml gel, dissolve 1.5 g agarose in 100 ml 0.5× TBE. Add 1 μl of Gel Red stain per 10 ml, e.g., in the above example add 10 μl of Gel Red stain.
3. Heat in a microwave on a medium-high power setting until all the agarose has dissolved.
4. Pour into a gel cast that has been sealed at both ends, add a comb (a comb producing 1 mm length wells will be sufficient). Leave to set (*see* **Note 7**).
5. Once the gel has set, unseal the ends, remove the comb, and place in a gel tank containing 0.5× TBE.
6. For each PCR prepared in Subheading 3.1, load 8 μl onto the gel in individual wells. In a separate well, add 5 μl of DNA size ladder.
7. Run the gel at 100 V for 45 min with an unlimited current.
8. Visualize the gel under UV light. The DNA size ladder should be sharp, well separated, and bright against the background of the gel. For the entire control region PCR, a visible band at about 1100 bp indicates successful amplification while if the shorter amplicon protocol has been followed then bands should be visible around the 300 bp mark (there should be three bands present for the first Midi-Mito reaction and two present for the second Midi-Mito reaction; however, due to the similarity in amplicon size between two of the products in the first Midi reaction it is probable that only two bands will be easily differentiated).
9. Should there be bands in the negative control reactions matching the size expected for the product then all relevant PCR products should be disposed of and the protocol restarted from the PCR step.
10. Providing the PCR positive has worked, should it be evident that PCR has failed for one or more samples then processing of those particular failed PCRs should stop and the PCR stage be repeated using the Midi-Mito strategy for those samples.

3.3 PCR Cleanup

1 For each reaction shown to have viable amplified products in the previous step, add 1.5 μl of ExoStar 1-Step to the remaining 10–12 μl of PCR product (*see* **Note 8**). Additionally process the negative PCR control in the same way.

2 In a thermal cycler, incubate at 37 °C for 45 min followed by 80 °C for 15 min.

3.4 Sequencing

Sequencing reactions can be undertaken using PCR tubes; however, it is more convenient to use a semiskirted 96-well plate that can be directly loaded onto a capillary electrophoresis machine following the sequencing cleanup stage.

1. Separate sequencing reactions in the forward and reverse direction must be setup for every sample. A single primer is included in each sequencing reaction, and hence multiple sequencing reactions are needed for every PCR amplicon. The sequencing primers used depend on the PCR strategy undertaken, such that:
   (a) Full control region strategy—seven sequencing reactions are carried out on each single 1.1 kb PCR product using primers L15977, H16175, H16401, L16450, H274, L314, and H599.
   (b) Midi-Mito strategy:
   - Midi-Mito PCR A—six sequencing reactions are carried out using primers L15989, H16248, L16450, H180, L317, and H619.
   - Midi-Mito PCR B—four sequencing reactions are carried out using primers L16197, H16509, L109, and H389.
2. Sequencing reactions should be setup in accordance with the components and volumes detailed in Table 4 using the primers specified above such that there are seven sequencing reactions for each sample amplified using the full control region strategy and ten sequencing reactions for each sample amplified using the Midi-Mito strategy. Sequencing reactions are carried out in a total reaction volume of 6.8 μl.
3. Multiple master mixes can be prepared and aliquoted into the plate for components that are identical between multiple wells, e.g., a different master mix could be prepared for each sequenc-

Table 4

Sequencing reaction reagent concentrations and volumes

Reagent	Amount added per reaction (μL)
ABI Prism® BigDye™ terminator cycle sequencing v3.1 ready reaction kit	0.68
ABI Prism® BigDye™ terminator cycle sequencing v3.1 dilution buffer	1.02
Sequencing primer (5 μM)	0.408
Water	3.6
Purified PCR product	1
Total	6.8

ing primer such that each master mix contains the BigDye ready reaction mix, BigDye dilution buffer, a single specific sequencing primer, and water.

4. By default, 1 μl of purified PCR product is added to the sequencing reaction. Should the gel electrophoresis stage have shown poor amplification (weak bands) then additional purified PCR product can be added to the reaction at the expense of the water. Purified PCR products should not be brought into a pre-PCR area.
5. The positive and negative controls from the PCR stage should be taken the whole the way through the process, and hence the purified PCR products for the positive and negative PCR controls should be sequenced now alongside the rest of the samples.
6. Seal the plate with either PCR caps or a PCR seal.
7. Sequencing should be carried out in a PCR machine with a heated lid using the following program: an initial activation of 96 °C for 4 min followed by 25 cycles of 96 °C for 15 s, 50 °C for 10 s, and 60 °C for 2 min.

3.5 Sequencing Cleanup and Electrophoresis

1. Add 1.7 μl of 125 mM EDTA to each amplified sequencing reaction in the plate. Vortex the plate.
2. Add 20 μl of 100 % ethanol into the same wells now containing sequencing product and EDTA (*see* **Note 9**).
3. Vortex briefly then leaves at room temperature for a minimum of 15 min and a maximum of 24 h to precipitate sequencing products.
4. Centrifuge the plate at 2250 × *g* for 30 min.
5. Remove all supernatant by unsealing the plate, inverting it onto a wad of folded paper towels and then centrifuging in this inverted orientation for 30 s at 400 rpm.
6. Add 120 μl of 70 % ethanol, reseal the plate, and spin in a normal orientation for 10 min at maximum speed (or 4000 rpm if this is lower) (*see* **Note 10**).
7. Repeat **step 5** to remove the supernatant.
8. Leave the plate at room temperature until all residual ethanol has evaporated (under normal circumstances, 5 min should suffice).
9. Add 10 μl of deionized formamide to each well containing sequencing product.
10. Heat at 95 °C for 3 min and then snap-cool (for example, by putting the plate in ice).
11. The sequencing products can now be run on a capillary electrophoresis instrument such as the 3130xl or 3500 (both Life Technologies).

3.6 Sequence Analysis

The mitochondrial control region sequence obtained from each sample needs to be compared to the revised Cambridge Reference Sequence (rCRS) and any differences between the reference sequence and the sample noted. To accomplish this, the multiple sequences produced for each sample from the different sequencing reactions must first be aligned to the reference sequence. There are various software solutions that will facilitate this alignment, such as SeqScape (Life Technologies) or Sequencer (Gene Codes). Once this alignment is complete, there are some important points to note.

1. It is ideal to have sequence coverage in both the forward and reverse direction for the entire control region, although this may be difficult to achieve in some circumstances, such as following homopolymeric stretches.
2. In almost all samples, there will be a difference to the reference sequence at base 263 (A > G) and an insertion of a C base following nucleotide 315. Full details regarding the nomenclature to use for reporting such changes can be found in the International Society for Forensic Genetics' guidelines [13–17].
3. The mitochondrial control region sequence of the 2800 M positive control has changes with respect to the rCRS at the following positions: 16519 (T > C), 152 (T > C), 263 (A > G), 315.1C, 477 (T > C).
4. Length heteroplasmy (the presence of multiple similar sequences that differ only in length due to the insertion or deletion of a base) is a common occurrence in mitochondrial analysis and is especially problematic in three regions:
   (a) If there is T > C change at base 16189 then this creates a poly-C stretch between bases 16184 and 16193 and almost inevitably results in multiple mitochondrial sequences with differing numbers of C nucleotides in this stretch. The consequence of this is that any sequencing reactions covering this area will be unreadable following the poly-C stretch because there will be multiple different sequences (offset by a few bases) displayed concurrently.
   (b) There are two more poly-C stretches between 303–309 and 311–315, and it is possible to observed length heteroplasmy here as well.
   (c) There is another poly-C stretch between bases 568 and 573, and in some individuals, there can be length heteroplasmy following this stretch of C nucleotides if additional C bases have been inserted into the sequence. This can result in failure of the H599 or H619 sequences.

5. Point heteroplasmy (two different bases at a single position) can also be observed, albeit at a lower frequency than length heteroplasmy. More detail on heteroplasmy can be found in the literature [17].
6. It is possible to observe mitochondrial sequence differences (usually at a single nucleotide position) between different tissues in an individual, and even between different hairs from the same individual. This usually manifests as point heteroplasmy but can be seen as a complete base change.
7. Using a web-based portal called EMPOP (www.empop.org), it is possible to obtain a frequency for the generated sequence (*see* **Note 11**).
8. Phylogenetic analysis can help to identify any errors in the final result; however, this can be difficult to undertake manually using phylotree (http://www.phylotree.org/tree/main.htm) unless relatively experienced. Alternatively, when examining the results from EMPOP, the program will display all those sequences that match closely, but not exactly, to your queried sequence. Here, for example, it may highlight that most similar sequences have a change at position 489 while your queried sequence has a change at 498, raising the possibility of a transcription error with the last two digits.

4 Notes

1. The DNA extraction method chosen will depend on the substrate, for example, specialized extraction methods are required for hair or bone. Normal DNA quantification methods are not very useful due to the types of samples used—total DNA quantification may be predominantly comprised of bacterial DNA while human specific methods target only the nuclear DNA whose concentration may bear no relation to the mitochondrial DNA concentration.
2. We notice no significant base misincorporation when using the MyTaq HS premix; however, it is also possible to use a proofreading polymerase such as Advantage HF 2 (Clontech).
3. It is possible to use any DNA sample as a positive control if it contains a consistent level of mitochondrial DNA and the mitochondrial sequence is known. 2800 M is stated here in order to provide a known set of mutations that can be referred to when undertaking analysis for the first time.
4. This is a nontoxic, although more expensive, alternative to using ethidium bromide to visualize the DNA bands. Ethidium bromide staining would also be acceptable, as would other nontoxic dyes.

5. Any characterized positive control can be used.
6. It is recommended to perform the PCR quality check stage (using gel electrophoresis) in order to gauge whether there is sufficient product to proceed with sequencing and estimate how much product to use in the sequencing reaction; however, it is possible to skip this stage and move straight to the sequencing stage entering an arbitrary amount of PCR amplicon into the sequencing reaction.
7. The time taken for the gel to set will depend on a number of factors including the ambient temperature and thickness of the gel; however, most gels will set within 20 min.
8. The PCR cleanup is performed to remove any unincorporated dNTPs and primers left over from the PCR stage. The method employed here utilizes two enzymes to accomplish this task; however, it is alternatively possible to achieve this end using a column cleanup approach.
9. The EDTA must be added to the amplified sequencing reaction before the 100 % ethanol because any direct contact between the EDTA and ethanol will result in precipitation of the EDTA—if a white precipitate/deposit is observed in any wells then either this EDTA/ethanol step was performed incorrectly or the EDTA used was too concentrated.
10. Care should be taken not to disturb the pellet when adding the 70 % ethanol, and the plate should not be vortexed after the ethanol has been added. The use of a 70 % ethanol concentration is important to keep the sequencing products precipitated, and hence 70 % ethanol should be made up fresh from absolute ethanol if the previous dilution is over 2 weeks old.
11. Reports should state the sequence range analyzed (e.g., 16,024–576) and all changes from the rCRS. Sequence frequency estimations can be obtained from EMPOP along with a 95 % confidence interval, for example, the 2800 M control region sequence is currently observed ten times out of 26,127 samples giving a frequency of 0.00038 with a 95 % confidence interval for this frequency between 0.00018 and 0.00070. Depending on the particular case circumstances it may be appropriate to report the frequency within the entire database, or more typically report the frequencies in one or more metapopulations (e.g., African, Westeurasian, and East Asian).

References

1. Anderson S, Bankier AT, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465
2. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA [5]. Nat Genet 23:147
3. Satoh M, Kuroiwa T (1991) Organization of multiple nucleoids and DNA molecules in

mitochondria of a human cell. Exp Cell Res 196:137–140

4. Robin ED, Wong R (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol 136:507–513
5. Torroni A, D'Urbano L, Rengo C, Scozzari R, Sbracia M, Manna C et al (1998) Intracytoplasmic injection of spermatozoa does not appear to alter the mode of mitocondrial DNA inheritance. Hum Reprod 13:1747–1749
6. Jenuth JP, Peterson AC, Fu K, Shoubridge EA (1996) Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. Nat Genet 14:146–151
7. Foran DR (2006) Relative degradation of nuclear and mitochondrial DNA: an experimental approach. J Forensic Sci 51:766–770
8. Bandelt H-J, Kong Q, Parson W, Salas A (2005) More evidence for non-maternal inheritance of mitochondrial DNA? J Med Genet 42:957–960
9. Lyons EA, Scheible MK, Sturk-Andreaggi K, Irwin JA, Just RS (2013) A high-throughput Sanger strategy for human mitochondrial genome sequencing. BMC Genomics 14:881
10. Eichmann C, Parson W (2008) '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 122:385–388
11. Ward RH, Frazier BL, Dew-Jager K, Pääbo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci U S A 88:8720–8724
12. Brandstätter A, Peterson CT, Irwin JA, Mpoke S, Koech DK, Parson W et al (2004) Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. Int J Legal Med 118:294–306
13. Parson W, Bandelt H (2007) Extended guidelines for mtDNA typing of population data in forensic science. Forensic Sci Int Genet 1:13–19
14. Tully G, Bär W, Brinkmann B, Carracedo A, Gill P, Morling N et al (2001) Considerations by the european DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. Forensic Sci Int 124:83–91
15. Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Holland M et al (2000) DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. Int J Legal Med 113:193–196
16. Carracedo A, Bär W, Lincoln P, Mayr W, Morling N, Olaisen B et al (2000) DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int 110:79–85
17. Irwin JA, Saunier JL, Niederstätter H, Strouss KM, Sturk KA, Diegoli TM et al (2009) Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples. J Mol Evol 68:516–527

Chapter 13

Whole Human Mitochondrial DNA Sequencing

Igor V. Ovchinnikov, Mathew J. Malek, Katelyn Kjelland, and Kenneth Drees

Abstract

Mitochondrial DNA (mtDNA) can help in the identification of biological evidence recovered from crime scenes and human remains. Typically the hypervariable regions are targeted for sequencing; however, more discriminating profiles are obtained if the whole genome is sequenced. Different approaches exist as to how best amplify and sequence whole mtDNA from forensic specimens. Here, we describe a method based on two-round PCR, combining multiplex and simplex PCRs. This method has been used in the analysis of mitochondrial genomes from archival saliva samples applied to FTA® cards after 10 years of transportation and preservation, without special protection. It is expected that this technique can be also used for the analysis of other old biological specimens directly or with modifications related to the level of DNA degradation.

Key words Whole mtDNA, DNA sequencing, Multiplex PCR, Simplex PCR

1 Introduction

Mitochondrial DNA (mtDNA) is a trace evidence of biological origin used for identifying perpetrators. It is also valuable for the identification of human remains based on linkage to maternal relatives. Since the first publications on isolation and sequencing of mtDNA from a variety of biological specimens [1, 2], it became clear that this molecular marker can be of crucial significance in forensic identifications, particularly when samples with low amounts of DNA and/or degraded DNA are available.

The first application of mtDNA in forensics was based on the sequencing of hypervariable sequence regions I and II (HVS-I and HVS-II) [3]. These segments are relatively small (about 400 bp each) and accumulate far more nucleotide polymorphisms than other regions of the mitochondrial genome due to a comparatively high mutation rate. However, power of discrimination using solely this region can be low, as many people share identical sequences, otherwise known as haplotypes, in HVS-I and HVS-II.

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_13,

Additionally, it is sometimes difficult to correctly determine an individual's haplogroup because similar polymorphisms in HVS-I and HVS-II may found in different haplogroups, or the nucleotide polymorphisms are noninformative. To overcome this limitation of partial DNA analysis, the whole mitochondrial genome sequences can be amplified and sequenced. In the majority of cases, this allows for the unambiguous placement of a given mtDNA genome into a haplogroup.

There are two approaches to whole human mtDNA sequencing. The first traditional approach is based on amplification of overlapping segments covering the entire genome, followed by Sanger sequencing. The second method was recently developed, based on a next-generation sequencing technology. Because the latter approach is currently under evaluation and may not yet meet forensic standards [4, 5], this chapter focuses on the multiplex PCR and Sanger sequencing technique, which is achievable in the majority of forensic laboratories.

Numerous PCR-based protocols have been developed to sequence the whole human mtDNA. These protocols differ in the number of PCR primers encompassing amplicons of varying size or in the amount of sequencing primers annealing to large PCR fragments (up to approximately 8.5 kb) [6–11]. Procedures are optimal for different applications depending on the quality and quantity of biological specimens and DNA isolated from them. In general, whole mtDNA sequencing protocols applicable for forensics should pursue the following criteria developed in [11] and supplemented in this chapter:

1. Each nucleotide should be determined from at least two independent sequencing reactions using different primers [11].
2. The number of PCR amplicons or sequencing primers required to cover the entire mitochondrial genome should be kept minimal [11].
3. Primers should produce an optimal signal-to-noise ratio, allowing for heteroplasmy detection [11].
4. When analyzing degraded samples primers should encompass relatively short PCR segments to permit stable amplification of mtDNA amplicons from degraded DNA and samples with low amounts of DNA.

Old and partially degraded samples are frequently met in forensic investigations. However, no approach has been specifically designed for the reconstruction of whole mtDNA from such sampling. In this chapter, we demonstrate that the human whole mtDNA can be efficiently amplified and sequenced from saliva specimens collected in 1997 in Central Asia. These samples were placed in 1.5 ml tubes and transported at ambient temperature. Tubes were then stored in a freezer at –20 °C for 10 years before

applying the saliva samples to FTA cards. Some recent protocols suggest carrying out amplification of the whole mitochondrial genome with only two overlapping amplicons of about 8.5 kb each [10, 11]. This technique is useful for relatively fresh blood and biopsy samples, but it was unable to produce results for old saliva samples with degraded DNA. Such difficult samples require an approach based on multiplex PCR followed by simplex PCRs with specific primer pairs and DNA sequencing of purified amplicons. We found that the primers published in [7] are better suited for successful reconstruction of whole mtDNA sequences from archival saliva specimens. Primers used for the first-round and second-round PCRs are listed in Table 1.

2 Materials

2.1 DNA Extraction

1. FTA Purification Reagent (Whatman).
2. Puncher for making 1.2 mm or 2.0 mm FTA disks (Whatman).
3. Two 100 ml beakers.
4. 0.2 ml PCR tubes or 96-well PCR plates (Fisher Scientific, or any vendor).
5. 10 % bleach solution (Clorox).

2.2 PCR

1. GoTaq DNA Polymerase (Promega).
2. 5× Colorless GoTaq Flexi Buffer (Promega).
3. 5× Green GoTaq Flexi Buffer (Promega).
4. 25 mM $MgCl_2$ solution (Promega; lots of alternative are available).
5. dNTPs.
6. Assorted forward and reverse primers (Tables 1 and 2).
7. Nuclease-free water.
8. 0.2 ml PCR tubes or 96-well PCR plates.

2.3 Agarose Gel

1. Quick Dissolve Agarose (BioExpress) or identical agarose from any vendor.
2. 10× TAE buffer (Fisher Scientific, or any vendor).
3. Ethidium bromide solution at concentration 10 mg/ml (Fisher Scientific or any vendor; alternative stains as GelRed and GelGreen from Biotium are available).
4. 100 bp PCR DNA ladder (Fisher Scientific).
5. 6× Blue/orange loading dye (Promega).

2.4 PCR Purification

1. DNA Clean & Concentrator-5 capped columns with DNA Binding Buffer and DNA Wash Buffer (Zymo Research).

Table 1
Primer pairs used for the PCR of human mtDNA[a]

Forward primer	Primer sequence 5′ → 3′	5′ coordinate	Reverse primer	Primer sequence 5′ → 3′	5′ coordinate
First-round PCR primers:					
SC-A-F	ATCCTACCCAGCACACAC	503	SC-A-R	GATTTGCCGAGTTCCTTTTACT	2484
SC-B-F	CTGACAATTAACAGCCCAATATC	2364	SC-B-R	GAATGCTGGAGATTGTAATGGG	4249
SC-C-F	CCAACTCATACACCTCCTATG	4155	SC-C-R	GGTAAGAGTCAGAAGCTTATGT	6220
SC-D-F	AATACCCATCATAATCGGAGG	6113	SC-D-R	GAGTACTACTCGATTGTCAACG	8017
SC-E-F	GGCGGACTAATCTTCAACTC	7925	SC-E-R	GTGAAATATTAGTTGGCGGATG	9884
SC-F-F	CATTTCCGACGGCATCTA	9767	SC-F-R	GCTAGGCAGAATAGTAATGAGG	11748
SC-G-F	CATTGCATACTCTTCAATCAGC	11614	SC-G-R	TTGACCTGTTAGGGTGAGAAGA	13638
SC-H-F	ATCATACACAAACGCCTGAGC	13539	SC-H-R	CGTCTTTGATTGTGTAGTAAGGG	15431
SC-I-F	CCACCTCCTATTCTTGCACG	15331	SC-I-R	TGCTAAAGGTTAATCACTGCTG	836
Second-round PCR primers:					
1F	CACACACACCGCTGCTAAC	516	1R	GATATGAAGCACCGCCAGG	1190
2F	GAACACTACGAGCCACAGC	1138	2R	TCATCTTTCCCTTGCGGTAC	1801
3F	AATTGAAACCTGGCGCAATAG	1756	3R	TGAGCATGCCTGTGTTGGG	2444
4F	ACCAACAAGTCATTATTACCC	2395	4R	TGAACTCAGATCACGTAGGAC	3074
5F	GGATCAGGACATCCCGATG	2995	5R	AACGGCTAGGCTAGAGGTG	3645
6F	TAGCTCTCACCATCGCTC	3536	6R	GATTGTAATGGGTATGGAGAC	4239
7F	TCCTACCACTCACCCTAGC	4184	7R	GTCATGTGAGAAGAAGCA	4869
8F	CACCCCTCTGACATCCGG	4832	8R	AGTATTGCAACTTACTGAGG	5570

9F	AATACAGACCAAGAGCCTTC	5526	9R	GGGAAACGCCATATCGGG	6188
10F	TACCCATCATAATCGGAGGC	6115	10R	AATATATGGTGTGCTCACACG	6781
11F	CTATGATATCAATTGGCTTCC	6730	11R	GGCATCCATATAGTCACTCC	7398
12F	CCTAATAGTAGAAGAACCCTC	7349	12R	CTCGATTGTCAACGTCAAGG	8009
13F	ATTATTCCTAGAACCAGGCG	7960	13R	TGATGAGATATTTGGAGGTGG	8641
14F	ACAATCCTAGGCCTACCCG	8563	14R	GATAGGCATGTGATTGGTGG	9231
15F	AGCCTCTACCTGCACGAC	9181	15R	GGATGAAGCAGATAGTGAGG	9867
16F	ACTTCACGTCATTATTGGCTC	9821	16R	AGTGAGATGGTAAATGCTAG	10516
17F[b]	CTGAACCGAATTGGTATATAG	10394	17R	TCGTGATAGTGGTTCACTGG	11032
18F	ACAATCATGGCAAGCCAACG	10985	18R	TTATGAGAATGACTGCGCCG	11708
19F	AGCCACATAGCCCTCGTAG	11633	19R	TGGTTATAGTAGTGTGCATGG	12361
20F	CTATCCATTGGTCTTAGGC	12284	20R	TTTGCCTGCTGCTGCTAGG	13005
21F	CGCTAATCCAAGCCTCACC	12951	21R	TATTCGAGTGCTATAGGCGC	13614
22F	TTACTCTCATCGCTACCTCC	13568	22R	GGTTGATTCGGGAGGATCC	14276
23F	CCCATAATCATACAAAGCCC	14227	23R	GTTGAGGCGTCTGCTGAG	14928
24F	ACTACAAGAACACCAATGACC	14732	24R	TGTAGTAAGGGTGGAAGGTG	15419
25F	TAGGAATCACCTCCCATTCC	15372	25R	GTCAATACTTGGGTGGTACC	16067
D1F	AATGGGCCTGTCCTTGTAG	15879	D1R	AACGTGTGGGCTATTTAGGC	16545
D2F	CGACATCTGGTTCCTACTTC	16495	D2R	CTGGTTAGGCTGGTGTTAGG	389
D3F	CGCTTCTGGCCACAGCAC	315	D3R	GGTGTGGCTAGGCTAAGC	803

[a]The primer sequences are taken from [7]. The 5′ coordinates correspond to the positions at the revised Cambridge Reference Sequence of the human mtDNA [13].
[b]Primer 17F sits over nucleotide position 10400 characteristic for haplogroup M. A second primer 17.1F with 5′ coordinate 10356 (CTAAGTCTGGCCTATGAGTG) could be added to sequence through 10400.

Table 2
Appropriate primer mixing for the first-round PCRs

First-round PCR mixtures	Primer pairs mixed in first-round PCR	Primer pairs used in second-round (simplex) PCRs
1 (multiplex)	SC-A, SC-D, SC-F, SC-H	1, 2, 3, 10, 11, 12, 16, 17, 18, 22, 23, 24
2 (multiplex)	SC-C, SC-E, SC-G, SC-I	7, 8, 9, 13, 14, 15, 19, 20, 21, 25, D1, D2, D3
3 (simplex)	SC-B	4, 5, 6

2. ZR-96 DNA Clean-up Kit with DNA Binding Buffer, Wash Buffer, Silicon-A Plate, Collection Plate, and Elution Plate (Zymo Research).
3. Nuclease-free water.

2.5 DNA Sequencing

1. BigDye Terminator v.3.1 Cycle Sequencing Kit with 5× Dilution Buffer and the BigDye reaction mixture (Life Technologies).
2. Sequencing primers corresponding to the forward and reverse PCR primers.
3. Nuclease-free water.

2.6 DNA Sequencing Purification

1. BigDye XTerminator Purification Kit with XTerminator solution and SAM solution (Life Technologies).

3 Methods

At all times during the extraction and amplification of human mtDNA, rigorous precautions must be followed in order to minimize the possibility of cross-transfer of mtDNA molecules between different samples. Negative extraction and amplification controls should be included to detect such cross-transfer and contamination. Pre-PCR and post-PCR procedures are to be divided between separate, isolated lab rooms and performed inside DNA workstations specially dedicated for each technological step such DNA isolation, PCR setup, and preparing DNA sequencing reactions after at least 2 h of irradiation under UV light in the hood to be used. After amplification, PCR tubes should be opened and agarose gel electrophoresis runs in a specialized electrophoresis room. Lab coats, reagents, and supplies must not be brought to DNA isolation and pre-PCR room from post-PCR/electrophoresis room.

3.1 Puncher and Cutting Square Cleaning Procedure

1. Set out two 100 ml beakers. Place 15 ml of 10 % bleach solution into one of the beakers and 20 ml of deionized water into the other beaker.
2. Briefly dunk the puncher into the bleach and remove. Follow quickly with dunking the puncher into the water. Do not let the puncher sit in the liquids.
3. Click the end of the puncher onto a clean paper towel to rid the inside of any liquid.
4. Let puncher stand to air dry.
5. Pour a small amount of 10 % bleach solution onto the cutting square.
6. Wipe bleach across the cutting square with a clean paper towel. Be sure to wipe entire surface of the cutting square. Let stand to air dry.

3.2 DNA Extraction from FTA Discs

1. In a biosafety cabinet, remove FTA card from packaging. Unfold and lay on clean cutting square. Hold the card only on its edge, being sure to not touch the card directly.
2. Using a clean puncher for making 1.2 mm or 2.0 mm disks, press firmly with puncher on area where the specimen is distributed on card.
3. Place FTA disc in 0.2 ml PCR tube by clicking end of puncher to release the disc. Clean puncher and cutting square (see above) and repeat process (**steps 1–3**) until all specimens have been punched.
4. Finally, make a final card punch with a cleaned puncher in the area outside of the specimen area. Place this disc into a PCR tube labeled as a negative extraction control.
5. Add 200 μl of FTA purification reagent to each PCR tube.
6. Incubate at room temperature for 5 min, then remove and discard spent FTA reagent.
7. Repeat **steps 5** and **6** twice for a total of three washes with FTA reagent.
8. Add 200 μl nuclease-free water to each PCR tube.
9. Incubate for 5 min at room temperature, then remove and discard water.
10. Repeat **steps 8** and **9** once for total of two washes with water.
11. Dry samples at 56 °C in heat block or incubation chamber with caps open until the discs are dry (approximately 30 min).
12. Setup PCR as shown below; reaction will take place in same tube and will include FTA disc.

3.3 Polymerase Chain Reactions (PCRs)

3.3.1 First-Round (multiplex) PCR

1. In a PCR workstation, prepare the primer mixture by pipetting the pairs of forward and reverse primers (5 μM) together into one tube to obtain a primer solution in which each primer has a concentration of 1 μM. Primer pairs should be mixed in combinations according to Table 2 (*see* **Note 1**).
2. Prepare the multiplex PCR master mix as indicated in Table 3. After each master mix has been made, pipette 50 μl of master mix into each PCR tube (or the specific wells of a 96-well plate) that contain FTA disks, one PCR tube/well with a blank extraction control, and one empty PCR tube/well for a negative PCR control.
3. Place PCR tubes or plate in thermocycler using parameters specified for a particular thermostable DNA polymerase and PCR primers. Thermal cycling conditions optimal for GoTaq DNA Polymerase included an initial denaturation at 95 °C for 2 min, followed by 30 cycles of denaturation for 1 min at 95 °C, primer annealing for 1 min at 58 °C, and extension for 2 min at 72 °C. After the last cycle, the final extension for 5 min at 72 °C was added (*see* **Note 2**).

3.3.2 Second-Round (Simplex) PCR

1. In a PCR workstation after UV treatment lasting two hours, separately prepare the master mixtures for each primer pair used in second-round PCR according to the recipe indicated in Table 4.
2. After each master mix is made, pipette 20 μl of the appropriate master mix into PCR tubes or specific wells of a 96 well plate.
3. Add 5 μl of the appropriate template from the first-round PCR into the appropriate tube/well.
4. For every primer pair, pipette 20 μl of master mix into a new PCR tube or an empty well, adding 5 μl of nuclease-free water instead of DNA template. This will be your second-round PCR negative control.
5. Place PCR tubes or plate in thermocycler using parameters specified for a particular thermostable DNA polymerase and PCR primers. Thermal cycling conditions optimal for GoTaq DNA Polymerase included an initial denaturation at 95 °C for 2 min, followed by 30 cycles of denaturation for 1 min at 95 °C, primer annealing for 1 min at 58 °C, and extension for 1 min at 72 °C. After the last cycle the final extension for 5 min at 72 °C was added (*see* **Note 2**).

3.4 Agarose Gel Electrophoresis

1. In an electrophoresis room, prepare a 1 % agarose gel by adding 1 g agarose to 100 ml of 1× TAE buffer in a 250 ml Pyrex glass bottle. Melt the agarose in a microwave oven by heating the mixture for 1–2 min (depending on the oven's power) until dissolved, with the bottle's lid placed ajar on top to reduce buffer evaporation. Do NOT seal the lid on the bottle!

Table 3
First-round (multiplex) PCR recipe

Reagent	Volume per sample (μl)	Concentration in PCR
Nuclease-free water	Add to a final reaction volume of 50 μl	–
5× colorless GoTaq Flexi Buffer	10	1×
25 mM $MgCl_2$	3	1.5 mM
dNTPs, 5 mM each	2.5	0.25 mM each
Primer mix, 1 μM each primer	5	0.1 μM each primer
GoTaq DNA Polymerase, 5 units/μl	0.25	1.25 units
FTA disk (DNA template)	0	–

Table 4
Second-round (simplex) PCR recipe

Reagent	Volume per sample (μl)	Concentration in PCR
Nuclease-free water	Add to a final reaction volume of 25 μl	–
5× green GoTaq Flexi Buffer (*see* **Note 3**)	5	1×
25 mM $MgCl_2$	1.5	1.5 mM
dNTPs, 5 mM each	1.25	0.25 mM each
Forward primer, 5 μM	2.5	0.5 μM
Reverse primer, 5 μM	2.5	0.5 μM
Go Taq DNA Polymerase, 5 units/μl	0.25	1.25 units
First-round PCR product	5	–

2. Warning: the bottle will be hot! Carefully remove bottle and let cool on bench to approximately 65 °C. Add 1 μl of ethidium bromide with a concentration of 10 mg/ml. Swirl bottle 2–3 times to mix solution. Ethidium bromide is considered to be a hazardous chemical; practice caution in handling using protective wear, such as gloves and a lab coat. Properly dispose of ethidium bromide-containing gels and buffers according to your institution's biosafety regulation.
3. Pour into a gel tray with a comb corresponding to the number of your samples. Let the gel solidify for at least 30 min for complete polymerization.

4. Add 1 μl of a 100-bp DNA Ladder solution to the first well. Starting with the second well, pipette 3–5 μl of each second-round simplex PCR reaction and each negative control (FTA extraction, first-round, and second-round negative controls) into individual wells (*see* **Note 3**). Be sure each sample is loaded into the appropriate well and not drifting into the well next to it.
5. Run electrophoresis for approximately 30–60 min at 80–100 V, remove the gel from tray, and view with UV light using a UV transilluminator or gel documentation system (*see* **Note 4**). In the image, find DNA bands analogous to the expected size as determined from the positions of primers on the human mtDNA. Second-round PCR products should generate a single bright band. If the negative controls do not produce bright or faint PCR bands (Fig. 1), move on to purification of the simplex PCR mixture remaining in the corresponding PCR sample tubes or plate's wells (Subheading 3.5).
6. If DNA products of the expected length are present in the negative extraction and the first-round PCR negative controls, dispose of the tubes or wells with corresponding PCRs and repeat the first- and second-round PCRs for the particular primer pairs. If DNA bands of the expected size are present only in the second-round negative controls, discard the second-round PCR tubes or wells with the particular primer pair and setup new second-round PCRs with the same primer pair.

3.5 PCR Purification

We recommend using the Zymo-Spin Column Purification for PCR products in individual tubes and the Zymo Silicon-A Plate protocol when there are PCR samples in a plate to clean.

3.5.1 Zymo-Spin Column PCR Purification

1. Label a 1.5 ml reaction tube and spin column with specimen code.
2. Add 100 μl of DNA Binding Buffer to each PCR reaction tube.
3. Pipette PCR reaction mixture into Zymo-Spin Column in a collection tube and centrifuge at 13,000 rpm for 30 s.
4. Discard flow through and place column back into collection tube.
5. Add 200 μl Wash Buffer to the column and centrifuge at 14,549 × g for 30 s.
6. Discard flow through, place column back into collection tube and repeat **step 5** for a total of two washes.
7. Discard flow through and collection tube. Place column into appropriate 1.5 ml reaction tube from **step 1**.
8. Add 40 μl of water and let column stand for 1 min.
9. Centrifuge at 13,000 rpm for 30 s and discard spin column. Store products in a freezer.

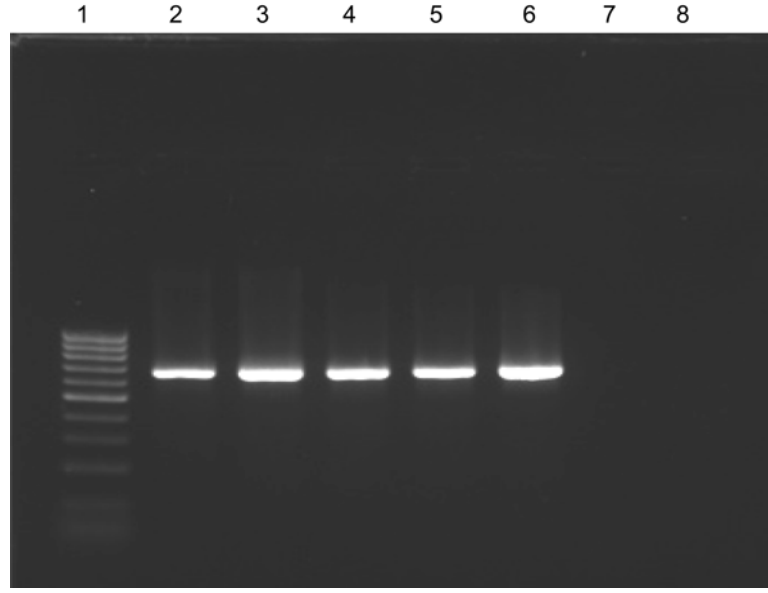


Fig. 1 Agarose gel electrophoresis showing PCR results. Five second-round PCR products obtained with the D1F and D1R primers are in *lanes 2–6*. The exACT-Gene ladder (Fisher BioReagents) is in *lane 1*. The negative extraction and PCR controls containing no PCR fragments are in *lanes 7* and *8*, respectively

3.5.2 Zymo Silicon-A Plate PCR Purification

1. If you carried out PCRs in the 96-well PCR plate, add 100 μl of DNA Binding Buffer to each DNA sample that is ready for clean up.
2. Transfer sample mixtures to wells of the Silicon-A Plate, which should be mounted on a collection plate.
3. Place specimen plate on scale and measure. Place balance plate on scale and add or remove water to balance both plates.
4. Centrifuge at 4,000 × *g* for 5 min until sample mixtures have been completely filtered. Discard flow through.
5. Add 300 μl of Wash Buffer to each well of the Silicon-A Plate. Rebalance the plates using the scale. Centrifuge at 4,000 × *g* for 5 min. Discard flow through. Repeat this step again for a total of two washes.
6. Add 40 μl of nuclease-free water directly to the column matrix in each well. Transfer the Silicon-A Plate onto an elution plate or a new 96-well plate, balance on the scale, and centrifuge at 4,000 × *g* for 5 min.
7. Products in the 96-well plate are ready for sequencing reactions. Label the plate and place in a fridge until sequencing.

3.6 DNA Sequencing

1. Using a NanoDrop Spectrophotometer, determine the concentration of PCR amplicons after purification of the PCRs. (The concentration of PCR amplicons can be also quantified based on comparison with a DNA ladder as 100 bp exACT-Gene DNA ladder, Fisher Scientific using an agarose gel.)
2. From Table 5 and the actual concentration of the PCR amplicons, estimate the volume of PCR fragments that should be added to the sequencing reactions.
3. Using the master mix protocol shown in Table 6, prepare the sequencing reactions with the correct amount of each reagent. Vortex and pipette the sequencing master mix into new 0.2 ml PCR tubes or in the appropriate wells of a 96-well plate.
4. Pipette correct volume of DNA template (PCR purification product) into each tube or well, ensuring the master mix being added contains the appropriate primer.
5. Place in thermal cycler and perform the sequencing reaction using an initial denaturation at 95 °C for 1 min, followed by 25 cycles of 10 s at 95 °C, 5 s at 53 °C, and 4 min at 60 °C. The temperature of primer annealing can be adjusted to the actual annealing temperature of a particular sequencing primer.

3.7 Sequencing Reaction Cleanup

1. After cycle sequencing, centrifuge the PCR tubes or the 96-well plate briefly, then pipette 20 μl of SAM Solution into every tube/well with the DNA sequencing products.
2. Vortex the XTerminator Solution until homogeneous. Using wide-bore pipette tips, pipette 5 μl of XTerminator Solution into each PCR tube/well.
3. Vortex for 30 min.
4. Centrifuge the PCR tubes or the 96-well plate briefly to separate out the XTerminator Solution.
5. The products can be analyzed on any Genetic Analyzer (Life Technologies) for Sanger sequencing (*see* **Note 5**). The reaction tubes or plate can also be kept in a fridge at 4 °C.

3.8 Data Analysis

1. Sequence each PCR fragment in both forward and reverse directions.

Table 5
Quantity of PCR amplicons of different size that are added to the sequencing reaction

Template (PCR amplicon size)	Quantity in the sequencing reaction
200–500 bp	3–10 ng
500–1000 bp	5–20 ng

Table 6
DNA sequencing reaction recipe[a]

Reagent	Volume per sample (μl)	Concentration in reaction
Nuclease-free water	Add to a final reaction volume of 6 μl	–
5× Dilution Buffer	1.5	1.25×
BigDye Reaction Mixture	1	–
Sequencing primer, 0.8 pmole	2	1.6 pmole
PCR product with concentration 3–20 ng/μl	1–1.5	4.5–20 ng

[a]One eighth of the cycle sequencing reaction mixture recommended by Life Technologies is used.

2. Open resulting electropherogram files and perform visual inspection of the mtDNA sequences by two researchers independently with BioEdit [12] or 4Peaks (http://www.nucleobytes.com). Many commercial softwares are available, but these are freeware that can be downloaded from the internet.
3. Align sample sequences to the revised Cambridge Reference Sequence [13] using MEGA5 [14] and identify variant sites in the sequences. MEGA5 can be replaced with any other software that can process DNA sequences including alignment, trimming, and assembling.
4. Submit mtDNA HVS1 and HVS2 polymorphisms to HaploGrep [15], an online application that associates mtDNA haplotypes with specific haplogroups by comparison to a phylogenetic tree of global human mtDNA variation (mtDNA tree Build 16 on PhyloTree.org) [16], to make a preliminary estimation of haplogroup association.
5. Verify the haplogroups determined through the HVS1 and HVS2 sequences by searching for the association of particular mutations found in your whole mtDNA sequence with the mtDNA tree Build 16 using the "Find" function in PhyloTree.org.

4 Notes

1. Do not amplify overlapping fragments of mtDNA in the same PCR, as it can lead to accumulation of short amplicons flanked by the reverse primer of the upstream fragment and the forward primer of the downstream fragment.
2. Each primer pair for the first-round and second-round PCRs was designed to anneal at 58 °C [7]. If different primers are used for the amplification of shorter amplicons, adjust annealing temperature according to the melting temperature for new

primers using a T_m (melting temperature) Calculator on the BioMath page of the Promega web site at www.promega.com/biomath. The annealing temperature of the primer pair equals the lowest T_m of any individual primer in the mixture, minus 3 °C. The time of extension for GoTaq DNA Polymerase is 1 kb/min. If a different thermostable polymerase is used, modify the cycling parameters according to the manufacturer's recommendations.

3. The Green and Colorless GoTaq Flexi Buffers give approximately equivalent yields. For PCRs containing the 5× Green GoTaq Flexi Buffer, load the reaction onto the agarose gel directly after amplification. Reactions containing the 5× Colorless GoTaq Flexi Buffer may also be loaded directly onto the agarose gel, but a tracking dye (e.g., 6× blue/orange loading dye) must be added to monitor the progress of electrophoresis [17]. Add a tracking dye to the DNA ladder solution.
4. UV transilluminator produces potentially harmful UV light. Protect eyes and skin from exposure to UV light. For safety, a gel documentation system has an automatic shut off for UV light when door is opened.
5. Each sequencing run results in a electropherogram file in .ab1 format.

Acknowledgments

The study was supported by the North Dakota EPSCoR and UND faculty start-up awards to I.V.O. The research protocol was approved by the University of North Dakota's Institutional Review Board (Protocol number IRB-201307-009) and the University of North Dakota's Institutional Biosafety Committee (Registration number IBC-201102-001).

References

1. Higuchi R, von Beroldingen CH, Sensabaugh GF, Erlich HA (1988) DNA typing from single hairs. Nature 332:543–546
2. Sullivan KM, Hopgood R, Lang B, Gill P (1991) Automated amplification and sequencing of human mitochondrial DNA. Electrophoresis 12:17–21
3. Sullivan KM, Hopgood R, Gill P (1992) Identification of human remains by amplification and automated sequencing of mitochondrial DNA. Int J Legal Med 105:83–86
4. Bandelt HJ, Salas A (2012) Current next generation sequencing technology may not meet forensic standards. Forensic Sci Int Genet 6:143–145
5. Parson W, Strobl C, Huber G, Zimmermann B, Gomes SM, Souto L et al (2013) Evaluation of next generation mtGenome sequencing using the Ion Torrent personal genome machine (PGM). Forensic Sci Int Genet 7:543–549
6. Rieder MJ, Taylor SL, Tobe VO, Nickerson DA (1998) Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. Nucleic Acids Res 26:967–973

7. Taylor RW, Taylor GA, Durham SE, Turnbull DM (2001) The determination of complete human mitochondrial DNA sequences in single cells: implications for the study of somatic mitochondrial DNA point mutations. Nucleic Acids Res 29:e74
8. Levin BC, Holland KA, Hancock DK, Coble M, Parsons TJ, Kienker LJ et al (2003) Comparison of the complete mtDNA genome sequences of human cell lines - HL-60 and GM 10742A - from individuals with pro-myelocytic leukemia and leber hereditary optic neuropathy, respectively, and the inclusion of HL-60 in the NIST human mitochondrial DNA standard reference material - SRM 2392-I. Mitochondrion 2:387–400
9. Coble MD, Just RS, O'Callaghan JE, Letmanyi IH, Peterson CT, Irwin JA et al (2004) Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. Int J Legal Med 118:137–146
10. Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA (2007) Whole-mtDNA genome sequence analysis of ancient African lineages. Mol Biol Evol 24:757–768
11. Fendt L, Zimmermann B, Daniaux M, Parson W (2009) Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences. BMC Genomics 10:139
12. Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98
13. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147
14. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739
15. Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G et al (2011) HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum Mutat 32:25–32
16. van Oven M, Kayser M (2008) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30:e386–e394
17. Glebs A, Stencel E, Knoche K (2003) The "Go-to" choice: GoTaq DNA polymerase. Promega Notes 83:21–24

Chapter 14

In-Solution Hybridization for the Targeted Enrichment of the Whole Mitochondrial Genome

B. Bekaert, R. Ellerington, L. Van den Abbeele, and R. Decorte

Abstract

A detailed protocol is presented for the targeted enrichment of whole mitochondrial genomes based on an in-solution hybridization strategy. Bait is produced in-house by sonication of two long-range PCR amplicons and ligation of biotinylated double-stranded adapters. Indexed target DNA is hybridized with the bait in a multiplex enrichment reaction and pulled down using magnetic streptavidin beads followed by subsequent post-enrichment PCR and sequencing on an Illumina MiSeq. This strategy removes the need for expensive commercial bait probes while allowing enrichment of multiple samples in a single hybridization reaction. The method is particularly suitable for degraded DNA as it is able to enrich short DNA fragments and is not susceptible to polymerase artifacts introduced during PCR-based assays.

Key words In-solution hybridization, Targeted enrichment, Mitochondrial DNA, Illumina MiSeq

1 Introduction

In most laboratories mitochondrial DNA (mtDNA) typing is still based on Sanger sequencing using a capillary electrophoresis-based platform. Usually this is limited to three polymorphic regions, called HVS-1, HVS-II, and HVS-III, which are located in the noncoding part of the mitochondrial genome despite the fact that more genetic information can be gained from the coding part of the mtDNA genome [1]. While the Sanger sequencing workflow is universally well established, it is laborious and has low throughput; therefore it is expensive to sequence larger regions such as the whole mtDNA genome. The advent of high-throughput sequencing (HTS) technologies has changed the field of genetics dramatically by providing greater throughput at reduced cost. In addition, Sanger sequencing can only provide a low resolution (10–20 %) on heteroplasmic variants while deep sequencing for variants can help investigate low frequency mutations in human disorders or interpret mixture samples for forensic purposes [2]. Despite advances in technology, whole genome sequencing is still very costly especially

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_14,

when only a part of the genome is of interest to the investigator. The enrichment of particular regions of the genome allows efficient sequencing of targets while omitting irrelevant sequences thereby allowing more samples to be multiplexed during a single run on a HTS instrument. Several techniques are available to enrich samples for specific genomic regions, with PCR being the most traditionally used technique. Published protocols for mtDNA sequencing with HTS instruments have used either singleplex or multiplex PCR-based enrichment methods followed by adaptor ligation [3–8]. This strategy, however, has some severe drawbacks: a high PCR cycle number for low quantity samples can lead to misincorporation of bases by DNA polymerases, PCR inhibition is a frequent occurrence especially for samples recovered from soil, short fragment sizes from degraded DNA require many short amplicons, and jumping PCR can create chimeric amplicons [9]. PCR has therefore lost popularity in favor of other strategies that do not require amplification of the target DNA. Capture methods use hybridization probes immobilized on a surface or in solution that bind to their complementary target sequences.

The tiling design of the bait probes makes sure that there are several fragments available that cover the target region for efficient enrichment. Commercial kits are available from several companies, but are expensive and difficult to customize. An in-house protocol for targeted sequencing by in-solution hybridization was developed by Maricic and colleagues for the simultaneous analysis of 46 whole mtDNA genomes [10]. We have updated this protocol to simplify the indexing of target samples by making use of a commercially available indexing kit. In short, the protocol uses bait produced by long-range PCR, sheared to fragments between 150 and 850 bp, and ligated to biotinylated adaptors, to enrich target DNA from a sonicated (in the case of modern DNA samples) and indexed input sample during a 48-h hybridization reaction. Magnetic streptavidin beads are used to capture the bait-target complex, and heat in combination with NaOH separates both DNA strands for subsequent sequencing after low cycle post-enrichment PCR.

2 Materials

2.1 Bait Preparation

1. DNA extracts.
2. PCR-grade water.
3. GoTaq® Long PCR kit (Promega).
4. Thermocycler.
5. Bioruptor plus (Diagenode).
6. Heat block.

7. Agarose gel electrophoresis equipment.
8. Quick-Load® 1 kb extended DNA ladder (NEB).
9. 25 bp DNA step ladder (Promega).
10. SYBRsafe (Life Technologies).
11. Agilent 2100 Bioanalyzer (Agilent Technologies).
12. High Sensitivity DNA analysis kit (Agilent Technologies).
13. Qubit® dsDNA HS Assay Kit (Life Technologies).
14. Qubit® (Life Technologies).
15. Quick Blunting™ Kit (NEB).
16. Quick Ligation™ Kit (NEB).
17. Primers (and blockers) (*see* Table 1).

2.2 Library Preparation

1. NEBNext® Ultra™ DNA Library Prep Kit (NEB).
2. NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1/Set 2) (NEB).
3. Optional: SYBRgold (Life Technologies).

2.3 Enrichment of Target

1. Incubator with rotator.
2. 20 % acetic acid.
3. Agilent gene expression hybridization kit (#5188-5242; Agilent).
4. Dynabeads® M-270 (#65305; Life Technologies).
5. Parafilm®.

2.4 Bait-Library-Enrichment

1. MinElute® PCR purification kit (Qiagen).
2. Agencourt® AMPure® XP beads (Beckman-Coulter).
3. 96–100 % ethanol.
4. Q5® Hot Start High-Fidelity DNA Polymerase (NEB).
5. Vortex.
6. Centrifuge.
7. Real-Time PCR machine.
8. Magnetic rack (Life Technologies).
9. 1.5 ml DNA LoBind Tubes (Eppendorf).
10. 0.2 ml tubes (Eppendorf).
11. 10× Oligo hybridization buffer (500 mM NaCl; 10 mM Tris–Cl, pH 8; 1 mM EDTA, pH).
12. BWT Buffer (10 mM Tris–HCl, pH 8; 1 mM EDTA; 1 M NaCl; 0.1 % Tween-20).
13. TET Buffer (10 mM Tris–Cl, pH 8; 1 mM EDTA, pH 8; 0.05 % Tween-20).

Table 1
Primers used for targeted enrichment of mtDNA

Long-range PCR primers	Sequence
BlockedA-F12186	C6Amino-GGCTTTCTCAACTTTTAAAGGATA
BlockedA-R3005	C6Amino-TGTCCTGATCCAACATCGAG
BlockedB-F2583	C6Amino-CCGTGCAAAGGTAGCATAATC
BlockedB-R12337	C6Amino-TTACTTTTATTTGGAGTTGCACCA
F12186	GGCTTTCTCAACTTTTAAAGGATA
R3005	TGTCCTGATCCAACATCGAG
F2583	CCGTGCAAAGGTAGCATAATC
R12337	TTACTTTTATTTGGAGTTGCACCA
Bait adaptors	Sequence
Primer-Bio-T	Biotin-TCAAGGACATCC*G
Primer B	CGGATGTCCTT*G
* indicates a phosphorothioate bond to prevent nuclease degradation of the oligos.	
Hybridization blockers	Sequence
BO1.P5.Forward	AATGATACGGCGACCACCGAGATCTACACT CTTTCCCTACACGACGCTCTTCCGATCT-phosphate
BO2.P5.Reverse	AGATCGGAAGAGCGTCGTGTAGGGAAAGAG TGTAGATCTCGGTGGTCGCCGTATCATT-phosphate
BO3.P7.PART1.Forward	AGATCGGAAGAGCACACGTCTGAACTCCAG TCAC-phosphate
BO4.P7.PART1.Reverse	GTGACTGGAGTTCAGACGTGTGCTCTTCC GATCT-phosphate
BO5.P7.PART2.Forward	ATCTCGTATGCCGTCTTCTGCTTG-phosphate
BO6.P7.PART2.Reverse	CAAGCAGAAGACGGCATACGAGAT-phosphate
Blockers for the index-adaptor region of the pooled DNA samples are used to prevent unspecific hybridization of sample fragments to one another. Blockers are phosphorylated at the 3′ end thus preventing their inadvertent use as primers in the amplification step.	
Enrichment PCR primers	Sequence
ImmortF	AATGATACGGCGACCACCG
ImmortR	CAAGCAGAAGACGGCATACGAGAT

Table 2
Buffer composition

10× Oligo hybridization buffer
500 mM NaCl
10 nM Tris–Cl, pH 8
1 mM EDTA, pH 8
HWT buffer
2.5 mM MgCl
50 mM KCl
15 mM Tris–HCl pH 8
0.1 % Tween-20
BWT buffer
10 mM Tris–HCl pH 8
1 mM EDTA
1 M NaCl
0.2 % Tween-20
TET buffer
10 mM Tris–Cl pH 8
1 mM EDTA pH 8
0.05 % Tween-20

14. 2× BWT Buffer (20 mM Tris–HCl, pH 8; 2 mM EDTA; 2 M NaCl; 0.1 % Tween-20).
15. HWT Buffer (2.5 mM MgCl; 50 mM KCl; 15 mM Tris–HCl, pH 8; 0.1 % Tween-20).
16. 125 mM NaOH.
17. Rox Reference dye (Life Technologies).

Primer sequences and buffers used in this protocol can be found in Tables 1 and 2 respectively. Treat all buffers with UV before use to prevent contamination.

3 Methods

3.1 Preparation of Bait

3.1.1 Long-Range PCR

1. Produce two long-range PCR products that together encompass the entire mitochondrial genome using the GoTaq® Long PCR kit in separate singleplex reactions. To avoid overrepresentation of overlapping sequences, blockers identical to the

primers with the exception of an amino group at the 5′ end are used (Table 1).

2. Each 25 µl reaction contains 2× GoTaq® Long PCR master mix, 1 µM of each forward and reverse primer, 1 µM of each blocker, and 10 ng DNA (*see* **Note 1**). Use the following PCR conditions: 95 °C initial denaturation 2 min, 30 cycles of 94 °C 15 s, 58 °C 30 s, and 65 °C for 9 min, final extension for 10 min at 72 °C. Confirm the amplification was effective by performing electrophoresis of the products on a 0.5 % agarose gel with Quick-Load® 1 kb extended DNA ladder. To visualize the DNA bands add 1 µl SYBR safe to every 10 ml of molten gel. Run the gel for 2–3 h at ~6 V/cm.
3. Purify the remaining product using the MinElute purification kit and quantify a 1/10 dilution using Qubit (dsDNA HS kit).
4. Pool both products in equimolar concentrations, and divide into aliquots of 1.5 µg of bait DNA in 100 µl by diluting with PCR-grade water.
5. Fragment the pooled sample using the Bioruptor plus sonication system from Diagenode. Set at (H)igh for 10 cycles, 30 s on, 90 s off to produce fragments of 300–500 bp length (*see* **Note 2**). Briefly spin down the sample every 3–4 cycles.
6. Confirm fragmentation was effective by determining the mean fragment size using a Bioanalyzer HS chip (*see* **Note 3**).

3.1.2 Annealing of Biotinylated Adapters

1. Each annealing reaction contains 50 µM of each oligo (Primer-Bio-T and Primer B) (*see* Table 1) and 10× Hybridization buffer in a 30 µl reaction.
2. Heat the mixture to 95 °C for 5 s and then ramp to 15 °C at a rate of –0.1 °C per s.
3. Dilute the reaction mixture with 210 µl of PCR-grade water.
4. At this point the mixture can be frozen at –20 °C until the next steps are carried out.

3.1.3 Blunt-End Repair of Bait

1. Perform blunt-end repair on the bait fragments using the NEB Quick Blunting™ Kit with 76 µl sheared mtDNA bait, 10× Blunting buffer, 4 µl Blunting enzyme mix, and 0.1 mM of dNTP mix for a total reaction volume of 100 µl. Alternatively, the End Repair module (NEB) can be used.
2. Incubate for 30 min at room temperature, purify using 1.6× AMPure XP beads, and elute in 15 µl of PCR-grade water.

3.1.4 Ligation of Adaptors to Bait

1. Ligate 1.25 µM of annealed adapters Bio-T/B to 15 µl of blunt-end repaired bait using 2× Quick ligase buffer and 4 µl of Quick ligase (Quick Ligation™ Kit) in a total reaction of 40 µl.

2. Incubate the reaction at room temperature for 15 min, purify the product using 1.6× AMPure XP beads, and elute in a final volume of 50 µl PCR-grade water.
3. Quantify a 1/10 dilution of the bait using Qubit (dsDNA HS kit).
4. Keep at –20 °C until needed in the "Enrichment" step.

3.2 Library Preparation

3.2.1 Fragmentation

1. Dilute 10 ng to 2 µg of target DNA in 100 µl of PCR-grade water.
2. Use the Bioruptor plus sonication system on the (H)igh setting for 10 cycles 30 s on, 90 s off, stopping every 3–4 cycles to briefly spin down the sample to produce fragments between 150 and 800 bp in length. The insert size required during size selection is dependent on the cycle sequencing kit used.
3. Use Bioanalyzer (HS chip) to confirm fragmentation was successful and to check that fragment lengths are between 150 and 800 bp.

3.2.2 End Repair, Ligation of Adaptors, and Indexing

The NEBNext® Ultra™ DNA Library Prep Kit is used for indexing of the samples. This kit has an input range for starting material between 5 ng and 1 µg of fragmented DNA. Up to 24 different samples can be indexed using the NEBNext® Multiplex Oligos Set 1 and Set 2 for Illumina (*see* **Note 4**).

1. Add 55.5 µl of fragmented DNA to 3 µl of End Prep Enzyme Mix and 10× End Repair Reaction Buffer in a final reaction volume of 65 µl. Incubate at 20 °C for 30 min, 65 °C for 30 min, and hold at 4 °C.
2. Add 15 µl of Blunt/TA ligase Master Mix, 2.5 µl of NEBNext® Adaptor for Illumina, and 1 µl of Ligation Enhancer for a total reaction volume of 83.5 µl.
3. Incubate this reaction mixture for 15 min at 20 °C.
4. Add 3 µl of USER™ enzyme, mix well, and incubate at 37 °C for 15 min.
5. Purify the sample with AMPure® XP beads using an adjusted bead/insert ratio according to the manufacturer's instructions. Elute in 28 µl of 0.1× TE buffer.
6. For indexing, add 25 µl of NEBNext® High Fidelity 2× PCR Master Mix to 23 µl of adaptor ligated target DNA, 1 µl of index primer, and 1 µl of universal primer (NEBNext® Multiplex Oligos Set 1/2). Use the following PCR conditions: 98 °C for initial denaturation and between 6 and 15 cycles of 98 °C for 10 s, 65 °C for 30 s, and 72 °C for 30 s. Finale extension of 5 min at 72 °C (*see* **Note 5**).
7. Freeze the indexed target at –20 °C until needed.

8. Once the indices are added, purify the samples using 1.6× AMPure XP beads and elute in 33 μl of 0.1× TE buffer and use Qubit to quantify each sample. Run a 1/5 dilution of an aliquot of the sample on a Bioanalyzer High Sensitivity chip to check the size distribution (*see* **Note 6**).

3.2.3 Optional (See Note 7)

1. Carry out a second "immortalizing" PCR to increase yield in separate aliquots ($n=4$ or 5) using primers ImmortF and ImmortR (*see* Table 1): 5× Q5 HF Buffer, 1.25 μM of primers ImmortF and ImmortR, SYBRgold (1/100.000), 0.25 μl ROX dye, and 2 μl indexed target with PCR conditions 98 °C for 30 s initial denaturation, 5–8 cycles of 98 °C 10 s, 65 °C 30 s, and 72 °C for 34 s, final extension for 5 min at 72 °C just before reaching the plateau phase.
2. Pool all 4–5 aliquots and purify using 1.6× AMPure XP. Elute in a final volume of 15 μl of 0.1× TE buffer.
3. Dilute an aliquot 1/5 for quantification using Qubit (HS dsDNA kit).

3.3 Enrichment

3.3.1 Bead-Bait Complex Preparation

1. Wash 5 μl of streptavidin-coated magnetic beads in 200 μl of BWT buffer and again in 200 μl of TET buffer.
2. To 500 ng of adapter-ligated bait, add an equal volume of 2× BWT buffer.
3. Heat the bait/BWT mixture for 1 min at 98 °C to denature the fragments and hold on ice until needed.
4. Add bait to the dry beads to resuspend.
5. Immobilize the biotinylated mtDNA fragments on the streptavidin-coated magnetic beads by rotating (or vortexing occasionally to ensure beads remain immobilized) for 20 min at room temperature.
6. Attract beads/bait to one end of the tube using a magnet and remove supernatant containing any un-reacted fragments, wash beads twice with 200 μl BWT which has been heated to 50 °C.
7. Resuspend the bait/beads in 50 μl of TET. (At this point the beads/bait can be stored at 4 °C until needed.)

3.3.2 Prepare Target DNA for Enrichment

1. While the beads/bait incubate mix the blockers with the pooled samples: 2 μg of pooled DNA, 2 μM of oligo's BO1.P5.F, BO1.P5.R, BO3.P7.PART1.Forward, BO4.P7.PART1.Reverse, BO5.P7.PART2.Forward and BO6.P7.PART2.Reverse, 10× Agilent blocking agent and 2× Agilent hybridization buffer (Agilent gene expression hybridization kit) and make with 0.1× TE to a total volume of 52 μl.
2. Incubate the target/blockers mixture to 95 °C for 3 min to denature fragments. Then cool to 37 °C for 30 min and spin down briefly.

3.3.3 Prepare Bait-Target Mixture

1. Use a magnetic rack to collect the beads from the bead/bait mix and remove them from the TET buffer.
2. Add the indexed DNA samples and blockers mix to the dry beads/bait complex.
3. Heat this mixture at 65 °C in a rotating oven for 48 h. Use Parafilm® to secure the lid.

3.3.4 After Hybridization

1. Incubate 200 μl of HWT buffer at 60 °C.
2. Remove hybridization mixture from the rotating oven.
3. Attract bait/beads to one end of the tube using a magnet and discard the supernatant containing any un-reacted fragments, wash beads with 200 μl BWT three times.
4. Add 200 μl of HWT buffer which has been preheated to 60 °C, incubate with beads at 60 °C for 2 min before collecting the beads magnetically and discarding supernatant. Repeat this once more.
5. Wash the beads with 200 μl BWT buffer.
6. Resuspend the beads in 100 μl TET buffer and remove the suspension into a new tube.
7. Remove the supernatant and resuspend the beads in 50 μl 125 mM NaOH and vortex for 10 s.
8. Attract the bait/beads to one end of the tube and remove suspension now containing the enriched DNA samples.
9. Add the NaOH supernatant to a neutralizing buffer of 500 μl PBI buffer (as outlined in the online MinElute handbook, add 120 μl of pH indicator to 30 ml of PB, both from MinElute PCR purification kit, to make PBI) and 3.8 μl 20 % acetic acid.
10. To the beads add 50 μl 125 mM NaOH that has been preheated to 95 °C. Heat the bead/NaOH mixture at 95 °C for 3 min. Remove suspension and add to the first suspension and PBI.
11. If the neutralizing buffer turns violet or orange (because it is alkaline), add more acetic acid, until the solution turns yellow (neutral) again.
12. Purify the two pooled supernatants and PBI using a MinElute column. Elute in a final concentration of 35 μl EB buffer.

3.3.5 Enrichment PCR

1. Amplify the enriched sample using QPCR with the Q5® High-Fidelity 5× Master Mix with 1.25 μM primers ImmortF and ImmortR, 2 μl ROX reference dye, 32 μl enriched DNA, and 4 μl 1:2000 diluted SYBRgold in a 200 μl reaction. Divide the reaction into 8 wells (25 μl each) of a QPCR plate and run with the following parameters: 94 °C for 2 min initial denaturation, 10–14 cycles of 95 °C 10 s, 60 °C 30 s, and 72 °C for

34 s, final extension for 10 min at 72 °C just before reaching the plateau phase. Alternatively, a standard PCR machine can be used for a 12 cycle amplification reaction in which ROX reference dye and SYBRgold can be left out of the reaction mixture.

2. Pool aliquots of the same sample and purify with 1.8× AMPure XP beads, elute in 20 μl PCR-grade water.
3. Quantify a 1/10 dilution of the enriched sample using Qubit (dsDNA HS kit) and run a HS Bioanalyzer chip to check the size distribution.
4. Dilute to a concentration of 2 nM.
5. Sequence the enriched sample using Illumina MiSeq using the standard protocol from the manufacturer with 1 % PhiX spike-in control (*see* **Note 8**).

4 Notes

1. Any human DNA sample can be used as bait.
2. No sonication is needed to shear the fragments to the appropriate size if the input DNA is degraded or low-quality material is used. This protocol has successfully been used for the simultaneous analysis of ancient DNA samples (A.D. 300–400; no sonication) and pristine DNA samples (with sonication).
3. Analyze the samples on the Bioanalyzer according to the manufacturer's instructions.
4. Previous studies have hybridized and sequenced up to 46 samples simultaneously using custom primers compatible with Illumina's chemistry [10].
5. The indexing PCR can be performed either on a regular PCR machine or on a real-time PCR machine. When using a real-time PCR machine, remove the sample before the stationary phase begins and immediately incubate for 5 min at 72 °C on a preheated regular PCR machine. Add SYBRgold (1/100.000) and 0.25 μl ROX dye to each 25 μl RT-PCR reaction. Size selection of the target inserts and enriched material can be carried out using the AMPure XP magnetic beads according to the instructions of the NEBNext® Ultra™ DNA Library Prep Kit.
6. The size range of the indexed product depends on the insert size but you should see a single broad peak in the range of 300–800 bp.
7. This step can be used to store indexed library material for secondary testing if needed. But beware: increasing PCR

cycles might produce unwanted duplicates of the indexed molecules and might introduce additional PCR artifacts by polymerase errors.

8. Follow the standard instructions from Illumina to prepare the run. The Illumina-specific adaptors can be used on the HiSeq instrument as well. When sequencing on other platforms the adaptors should be changed accordingly.

References

1. Parsons TJ, Coble MD (2001) Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. Croat Med J 42:304–309
2. Zhang C, Huang VH, Simon M, Sharma LK, Fan W, Haas R et al (2012) Heteroplasmic mutations of the mitochondrial genome cause paradoxical effects on mitochondrial functions. FASEB J 26:4914–4924
3. Mikkelsen M, Frank-Hansen R, Hansen AJ, Morling N (2014) Massively parallel pyrosequencing of the mitochondrial genome with the 454 methodology in forensic genetics. Forensic Sci Int Genet 12:30–37
4. Bekaert B, Massoli C, Anandarajah A, Van de Voorde W, Decorte R (2013) Multiplex DNA amplification and barcoding in a single reaction for 454 Roche sequencing: a comprehensive study on the control region of the mitochondrial genome. Forensic Sci Int Genet 4:e111–e112
5. McElhoe JA, Holland MM, Makova KD, Su MS-W, Paul IM, Baker CH et al (2014) Development and assessment of an optimized next-generation DNA sequencing approach for the mtGenome using the Illumina MiSeq. Forensic Sci Int Genet 13:20–29
6. Hamoy IG, Ribeiro-dos-Santos AM, Alvarez L, Barbosa S, Silva A, Santos S et al (2014) A protocol for mtGenome analysis on large sample numbers. Bioinform Biol Insights 8:127–134
7. Templeton JEL, Brotherton PM, Llamas B, Soubrier J, Haak W, Cooper A et al (2013) DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. Investig Genet 4:26
8. Parson W, Strobl C, Huber G, Zimmermann B, Gomes SM, Souto L et al (2013) Evaluation of next generation mtGenome sequencing using the ion torrent personal genome machine (PGM). Forensic Sci Int Genet 7:543–549
9. Brakenhoff RH, Schoenmakers JGG, Lubsen NH (1991) Chimeric cDNA clones: a novel PCR artifact. Nucleic Acids Res 19:1949
10. Maricic T, Whitten M, Pääbo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS One 5:e14004

Chapter 15

Enhanced DNA Profiling of the Semen Donor in Late Reported Sexual Assaults: Use of Y-Chromosome-Targeted Pre-amplification and Next Generation Y-STR Amplification Systems

Erin K. Hanson and Jack Ballantyne

Abstract

In some cases of sexual assault the victim may not report the assault for several days after the incident due to various factors. The ability to obtain an autosomal STR profile of the semen donor from a living victim rapidly diminishes as the post-coital interval is extended due to the presence of only a small amount of male DNA amidst an overwhelming amount of female DNA. Previously, we have utilized various technological tools to overcome the limitations of male DNA profiling in extended interval post-coital samples including the use of Y-chromosome STR profiling, cervical sample, and post-PCR purification permitting the recovery of Y-STR profiles of the male DNA from samples collected 5–6 days after intercourse. Despite this success, the reproductive biology literature reports the presence of spermatozoa in the human cervix up to 7–10 days post-coitus. Therefore, novel and improved methods for recovery of male profiles in extended interval post-coital samples were required. Here, we describe enhanced strategies, including Y-chromosome-targeted pre-amplification and next generation Y-STR amplification kits, that have resulted in the ability to obtain probative male profiles from samples collected 6–9 days after intercourse.

Key words Extended interval post-coital samples, Sexual assault evidence, Y-chromosome short tandem repeat (STR) analysis, Y-chromosome-targeted pre-amplification, Next generation Y-STR amplification kit, Forensic science

1 Introduction

For a variety of reasons, some victims of sexual assault provide vaginal samples more than 36–48 h after the incident. In these cases, the ability to obtain an autosomal STR profile of the semen donor from the living victim diminishes rapidly as the post-coital interval is extended [1]. Autosomal STR profiles (the gold standard for current DNA profiling) of the male donor are frequently not obtained from vaginal samples taken 24–36 h after intercourse due to technological impediments and limitations, such as analytical detection sensitivity limits or the "masking" of male DNA profile

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_15,

due to the overwhelming amount of female DNA present in these samples [1, 2]. While the ability to recover a male autosomal STR may not be possible, this failure is often not due to a complete absence of male cells (sperm and/or male epithelial cells). Classical forensic serology studies have shown consistently that spermatozoa (although few in number) persist in the vaginal canal three or more days after intercourse, and the reproductive biology literature also demonstrates that spermatozoa can be detected in the human cervix up to 7–10 days post-coitus [3–9]. Therefore, it is critical that ultrasensitive profiling strategies be developed for the analysis of the few sperm recoverable from extended interval post-coital samples to permit the recovery of critical probative information (i.e., DNA profiles of the male donor) for criminal investigations.

Y-chromosome STR (Y-STR) profiling has provided a valuable alternative to autosomal STR analysis for use with the analysis of sexual assault evidence as it targets only the male DNA present in admixed samples, even amongst an overwhelming amount of female DNA [2, 10–24]. We demonstrated the ability to utilize Y-STR profiling (commercially available Y-STR kits) for the analysis of post-coital samples collected 3–4 days after intercourse [2, 19]. While this was an improvement in the interval in which probative genetic information could be recovered from post-coital samples, we sought to further extend the interval in which a male Y-STR profile could be obtained. With the use of technical enhancements such as cervical sampling and post-PCR purification, we were able to obtain probative male profiles from samples collected 5–6 days after intercourse [2]. Despite this improvement, this was still well below the reported 7–10 day time frame reported in reproductive literature in which sperm were identified. We hypothesized that a pre-amplification of the small amount of male DNA present prior to standard Y-STR profiling would provide the necessary sensitivity increase to surpass the analytical detection limits of the existing Y-STR methods. We therefore developed a targeted genome pre-amplification system (Y-chromosome-targeted pre-amplification, Y-TPA) that specifically amplifies multiple STR loci on the Y-chromosome using a nested PCR approach [25, 26]. With nested PCR, first round amplification is performed with a single primer set for each target [27]. As with any standard PCR method, it is possible during this initial amplification to obtain unwanted amplification products (i.e., primer dimers and nonspecific amplification artifacts due to mis-priming), which can reduce the efficiency of amplification of intended targets. However with a nested PCR approach, the resulting first round amplification product is subjected to a secondary amplification with a second set of primers (nested primers, binding sites contained within the first round amplification products). If incorrect (i.e., nonspecific) amplification products were obtained in the first round amplification, they would not be amplified using the specific "nested" primers required

for the second round amplifications. This improved specificity therefore can increase sensitivity as the targets of interest are amplified more efficiently (i.e., enhancement of the desired targets, Y-STRs in this case). This improvement in specificity and sensitivity may not be observed with amplifications that, for example, simply increase the number of amplification cycles of a single reaction using single primer sets. Additionally, increased cycle number reactions can result in depletion of critical PCR reagents, thereby reducing the efficiency of the amplification in higher cycle numbers. This can result in little, if any, improvement in sensitivity. The developed Y-chromosome-specific pre-amplification method includes suitable first round amplification primers (Table 1) to allow for a multiplex pre-amplification of all 17 loci contained in the Yfiler® amplification system. It should be noted that 12 of these 17 loci are contained in the Promega Y-STR product, PowerPlex® Y System. The DYS389 outer or first round amplification primers encompass both products and therefore only a single amplification product is obtained after first round amplification unlike standard Y-STR typing using only the commercial multiplexes where two separate products are obtained. The results of our studies involving the use of Y-TPA demonstrate the ability to obtain male donor profiles in extended interval post-coital samples collected 6–9 days after intercourse [25, 26]. This represents a significant improvement in the time frame in which male profiles can be successfully recovered from post-coital samples.

While the results of our studies using Y-TPA clearly demonstrate that the use of enhanced Y-STR typing strategies is an effective strategy for the improved recovery of male DNA profiles from extended interval post-coital samples, operational crime laboratories may be reluctant to utilize such an approach in casework due to concerns regarding the increased risk of contamination due to an increase in cycle number and the need for additional sample manipulation during purification and secondary amplification steps. While these issues can be addressed with proper laboratory protocols, there is no commercial kit available for the pre-amplification multiplex, which is also not ideal for operational crime laboratories. Additionally, Y-TPA is limited to the recovery of 17-locus profiles, the number of loci routinely used during the time of development of this approach.

Since the time of development of the Y-TPA method, new "next generation" Y-STR kits have become available (PowerPlex® Y23, Promega and Yfiler® Plus, Life Technologies) which, apart from the incorporation of a more robust buffer chemistry, incorporate additional loci. The Promega PowerPlex® Y23 amplification kit contains 23 loci (six additional loci compared to Yfiler®), which could provide additional discriminatory potential particularly when partial profiles are obtained. The kit is reported to have a high sensitivity in the presence of female DNA (<0.1 ng, 1:6000 male:

Table 1
Primer sequences and concentrations for Y-chromosome-targeted pre-amplification (Y-TPA)

Locus	Final primer conc (μM)	Primer sequence (5′–3′)
DYS19	0.35	F: TGAAATCAAAAAATAATCACAGTCA R: AGGACTCAGGAATTTGCTGG
DYS385	0.15	F: CTAAAGTTCTACCCAAATTTTGTCAC R: GTAATTGTTTGAACCTGAAATGT
DYS389	0.20	F: AACAGCAGGTTCTCCGCTAC R: ACCTAATCTGTCAATGATTTTCTGTC
DYS390	0.60	F: AGTGTATCCGCCATGGTAGC R: GCAAGTGAACCAATAAATGAA
DYS391	0.06	F: CTAACCTATCATCCATCCTTATCTCTTG R: ATAGATGGATACATAAGACATTAGAATGG
DYS392	0.12	F: AGACAGCTGGTGTGCTCCTT R: TAGAGGCAGTCATCGCAGTG
DYS393	0.05	F: CGACCATGTGGCTGTGAGTC R: GCCAGATAACGTGTGTGGAA
DYS437	0.04	F: CAGCCTCAATTTCCTGGTCT R: TGCAGCCTGAGGAACAGA
DYS438	0.06	F: GAAGGAAATAGAGTAACACCATTAGAG R: ATCACCCAGGGTCTGGAGTT
DYS439	0.60	F: CGGTAGTTTCCTTTGCTGTA R: ATGCCTGGCTTGGAATTCTT
DYS448	0.13	F: CACCCGTGTAGGGAGATGTT R: CTCTTTCCTGAGTGGAGGTT
DYS456	0.10	F: CAGGTGCCAGTGCAACTAGA R: CCATCAACTCAGCCCAAAAC
DYS458	0.04	F: CATGGTGAAACATTGTCTCTTT R: TTCCTGCCACCACGC
DYS635	0.20	F: CTTAAACCCAGGAGGCAGAG R: GCGTGTCTGTATGTATATGAAAATGTG
H4	0.14	F: TGGTCAAAACACCATTTCCTC R: GGACAGAGTGGGTTCTGAAG

female DNA ratio) and a high tolerance for inhibitors. This "enhanced" kit, therefore, may be ideally suited for use with the extended interval post-coital samples in which trace amounts of male DNA are present admixed with an extremely high amount of female DNA. Alternatively, another next generation Y-STR system is available which includes both a new quantitation system (Quantifiler® Trio) and an amplification system (Yfiler® Plus) (Life

Technologies). The Quantifiler® Trio quantitation kit is designed as a fast and accurate decision tool, providing an indication of whether there is sufficient amounts of male DNA for analysis and also whether the sample should be processed using autosomal or Y-STR analysis (using the included M:F ratio). It has a shorter analysis time and increased sensitivity and accuracy compared to the company's previous products. The quantitation kit permits a determination of the amount of male DNA as well as total human DNA, and includes both a small and large autosomal target to provide an indication of sample quality (i.e., degradation index). The Yfiler® Plus amplification system is a 27-locus multiplex that contains all 17 loci from the original Yfiler® system plus ten new Y-STR loci including seven rapidly mutating Y-STR loci [28]. It is reported to have improved performance with challenging casework samples. Our results from the evaluation of these next generation Y-STR systems demonstrate the ability to obtain probative Y-STR profiles from sexual assault samples collected 4–9 days after intercourse [29].

With our successful use of Y-TPA and next generation Y-STR amplification systems for improved male profile recovery from extended interval post-coital samples, here we provide detailed protocols for these enhanced typing strategies. The experimental schema for the enhanced strategies is relatively straightforward: (1) isolation of male DNA from extended interval post-coital samples using a non-differential manual organic extraction with ethanol precipitation, (2) purification and isolation of the isolated male DNA, (3) simultaneous quantification of male and human DNA, and (4) amplification using Y-TPA/Yfiler® and/or amplification with a next generation Y-STR amplification system. We are hopeful that the incorporation of these methods into casework will invoke an expansion of the routinely used 72-h time frame for evidence collection in sexual assault cases. This should result not only in an increase in the conviction rate for such "extended post-coital interval" cases but also in exculpatory evidence that previously would have been missed with more traditional analysis.

2 Materials

2.1 General Equipment and Supplies

1. Pipets (0.5–10, 2–20, 20–200, 100–1000 μl).
2. Sterile, aerosol-resistant pipet tips.
3. Microcentrifuge tubes (0.2, 0.5, 1.5 ml).
4. Microcentrifuge tube racks.
5. Extraction tubes (1.5 ml).
6. Spin Ease Baskets.
7. Disposable transfer pipets.

8. Water bath.
9. Heat blocks.
10. Microcentrifuge.
11. Vortex.
12. Vacuum centrifuge.
13. Refrigerators, freezers.
14. Gloves.
15. 15 and 50 ml conical tubes.
16. Tweezers, scissors, and/or scalpels (single use).
17. Stir plates.
18. Stir bars.
19. pH meter.
20. Autoclave.
21. KimWipes®.

2.2 Manual Organic DNA Extraction (See Note 1)

1. DNA extraction buffer: 100 mM NaCl, 10 mM Tris–HCl pH 8.0, 25 mM EDTA, 0.5 % SDS. To prepare, combine 10 ml 5 M NaCl, 5 ml 1 M Tris–HCl pH 8.0, 25 ml 0.5 M EDTA pH 8.0, and 2.5 g SDS and bring volume to 500 ml with deionized water. Stir until all reagents are dissolved.
2. TE^{-4}: 10 mM Tris–HCl, 0.1 mM EDTA, pH 7.5. To prepare, dissolve 0.605 g Tris base in 400 ml deionized water. Adjust pH to 7.5. Add 0.0185 g EDTA disodium salt. Recheck pH and adjust to 7.5 if needed. Bring the final volume to 500 ml with deionized water. Autoclave.
3. Proteinase K (20 mg/ml).
4. 0.39 M dithiothreitol (DTT).
5. Ethanol, 70 %.
6. Ethanol, 100 %.
7. Phenol-chloroform isoamyl alcohol, pH 6.7 ± 0.2 (*see* **Note 2**).

2.3 DNA Extraction Concentration and Purification

1. MinElute PCR Purification kit (QIAGEN).
2. Sterile water (18.2 MΩ), pH 7.0–8.5.

2.4 DNA Quantitation (See Note 3)

1. Quantifiler® Trio DNA quantification kit (Life Technologies) (*see* **Note 4**).
2. MicroAmp® optical 96-well reaction plate (or equivalent).
3. MicroAmp® optical adhesive covers (or equivalent).
4. 7500 Real time PCR instrument (Life Technologies) (*see* **Note 5**).
5. Low TE buffer: 10 mM Tris, 0.1 mM EDTA, pH 8.0.

2.5 Y-Chromosome-Targeted Pre-amplification (Y-TPA)

1. Type-it® Microsatellite PCR Kit (QIAGEN).
2. Y-TPA primer mix: To prepare the mix, add the following primers at the specified final concentration (Table 1). Vortex well. The primer mix should be quality control tested with a male DNA control sample (15–25 pg input) prior to use.
3. Sterile water (18.2 MΩ).
4. MinElute PCR Purification kit (QIAGEN).
5. AmpFlSTR® Yfiler® Plus PCR Amplification kit (Life Technologies).
6. Thermal cycler.

2.6 Next Generation Commercial Y-STR Kit Amplification

1. AmpFlSTR® Yfiler® Plus PCR Amplification kit (Life Technologies).
2. PowerPlex® Y23 System (Promega).
3. Low TE buffer: 10 mM Tris, 0.1 mM EDTA, pH 8.0 (*see* **Note 6**).
4. Thermal cycler.

2.7 Capillary Electrophoresis (See Note 7)

1. Hi-Di™ Formamide (Life Technologies).
2. GeneScan™ 500 LIZ® Size Standard (Life Technologies).
3. GeneScan™ 600 LIZ® Size Standard (Life Technologies).
4. ABI Prism 3130 Genetic Analyzer (Life Technologies).
5. MicroAmp® optical 96-well reaction plate (or equivalent).
6. Plate septa, 96-well (Life Technologies).
7. Performance-optimized polymer POP-7 (Life Technologies).

3 Methods (See Note 8)

3.1 Post-coital Sample Collection

The recovery of a male DNA profile from the semen donor in extended interval (>3 days) post-coital samples requires the collection of samples from the cervical region of the victim. This is in addition to the low and/or mid-vaginal samples normally routinely recovered after a vaginal sexual assault. A minimum of two sterile cotton swabs should be taken from the cervix and/or fornix. One of the swab pair should be processed as described below and the other retained for possible re-analysis.

3.2 Manual Organic DNA Extraction

1. Properly label the appropriate number of 1.5 ml extraction tubes.
2. Remove the swab from the wooden stick or cut the wooden stick with sterile scissors just above the cotton portion so that the entire cotton part of the swab is used for extraction (*see* **Note 9**).
3. To each sample, add the following: 400 µl DNA extraction buffer, 13 µl proteinase K (20 mg/ml), 40 µl 0.39 M DTT (~10 % of extract volume, final concentration of 0.039 M).

4. Mix the samples on a vortex for ~2 s.
5. Incubate samples in a 56 °C water bath overnight.
6. Remove the swab pieces and place in a spin basket. Place the basket back into the original tube.
7. Centrifuge samples at 14,000 rpm (16,000 × *g*) for 5 min. Discard basket.
8. In a fume hood, add 400 μl of phenol/chloroform/isoamyl alcohol (amount equal to the extract volume). Mix well by inversion (solution should be milky in appearance). Do not vortex.
9. Centrifuge samples at 14,000 rpm (16,000 × *g*) for 5 min to separate the phases. The organic material will be trapped in the lower nonpolar layer, and the polar aqueous phase (top later) will contain the DNA.
10. Carefully remove the top aqueous layer and transfer to a new 1.5 ml tube. Be careful not to disturb the interface.
11. Add 1 ml cold absolute ethanol (100 %) to transferred aqueous layer. Mix by inversion.
12. Place samples at −20 °C for at least 1 h to precipitate DNA (may proceed overnight).
13. Centrifuge samples at 14,000 rpm (16,000 × *g*) for 15 min to pellet the DNA.
14. Remove the ethanol with a disposable transfer pipet without disturbing the pellet.
15. Wash pellet with 1 ml room temperature 70 % ethanol.
16. Centrifuge the samples at 14,000 rpm (16,000 × *g*) for 5 min.
17. Remove the ethanol with a pipet.
18. **Steps 15–17** can be repeated for multiple washes (2–4×) if desired.
19. Dry the pellet using a vacuum centrifuge for 10–15 min.
20. Add 75 μl TE-4 to each sample.
21. Place samples in a 56 °C water bath overnight to resolubilize the DNA.
22. DNA extracts can be stored at 4 °C for short-term storage or −20 °C for long-term storage.

3.3 DNA Extraction Concentration and Purification (MinElute) (See Notes 10 and 11)

1. Add ethanol to Buffer PE before use (see bottle label for volume).
2. Add 1:250 volume pH indicator to Buffer PB (referred to subsequently as Buffer PB-I; Buffer PB with pH indicator). The yellow color of the Buffer PB-I indicates a pH of ≤7.5.

3. Add 5 volumes of Buffer PB-I to 1 volume of the DNA extract (*see* **Note 12**). Ensure that the buffer color remains yellow (*see* **Note 13**).
4. Insert a MinElute column into a 2 ml collection tube (both provided in the kit). Add the sample/Buffer PB-I to the MinElute column.
5. Centrifuge columns at 13,000 rpm (14,800 × *g*) for 1 min. Discard flow-through and place the column back into the same collection tube.
6. Add 750 µl Buffer PE to the MinElute column and centrifuge columns at 13,000 rpm (14,800 × *g*) for 1 min.
7. Discard flow-through and place the MinElute column back into the same tube. Centrifuge the columns for an additional 1 min at 13,000 rpm (14,800 × *g*) to remove all residual ethanol.
8. Place the MinElute column in a clean 1.5 ml microcentrifuge tube.
9. Add 13 µl sterile water to the center of the column membrane (*see* **Note 14**).
10. Incubate the column at room temperature for 1 min.
11. Centrifuge the column 13,000 rpm (14,800 × *g*) for 1 min. Discard columns.
12. DNA extracts can be stored at 4 °C for short-term storage or −20 °C for long-term storage.

3.4 DNA Quantitation

1. Prior to running the quantitation, ensure that the real time instrument has been calibrated for dyes VIC®, ABY™, 6-FAM™, JUN™, and Mustang Purple™.
2. Prepare quantitation standards as listed in Table 2 (serial dilution). Ensure that each standard is vortexed well and briefly centrifuged.

Table 2
Preparation of Quantifiler® trio quantitation standards

Std	Conc (ng/µl)	Sample volume	Dilution buffer volume[a] (µl)
Std 1	20.0	20 µl 100 ng/µl stock[a]	80
Std 2	2.0	10 µl Std 1	90
Std 3	0.2	10 µl Std 2	90
Std 4	0.02	10 µl Std 3	90
Std 5	0.002	10 µl Std 4	90

[a]Provided in kit

3. Thaw Primer Mix and PCR Reaction Mix. Vortex and briefly centrifuge.
4. Prepare the PCR master mix (per sample) (prepare ~10 % extra to account for potential pipetting errors): 8 μl Quantifiler® Trio Primer Mix, 10 μl Quantifiler® Trio PCR Reaction mix. Vortex and briefly centrifuge.
5. Add 18 μl of the prepared master mix to the appropriate reaction wells.
6. Add 2 μl of sample, standard, or control to the appropriate wells. Use low TE buffer as a no-template control (negative control). Standards should be run in duplicate.
7. Seal the reaction plate with optical adhesive cover. Centrifuge the reaction plate at 3000 rpm for 30 s.
8. The cycling conditions are as follows: 95 °C, 2 min; followed by 40 cycles of 95 °C, 9 s; 60 °C, 30 s.
9. Analyze quantification results using HID Real Time PCR Analysis Software v. 1.1 (*see* **Note 15**).

3.5 Y-Chromosome-Targeted Pre-amplification (Y-TPA)

1. Label the appropriate number of 0.2 ml PCR tubes for the Y-TPA amplification.
2. Add up to 5 μl of concentrated DNA extract (maximum of 5 μl is used for samples with undetectable male DNA quantitation values).
3. Add appropriate amount of sterile water to bring reaction volume to 7.5 μl. Positive (male DNA sample of known profile, 15 pg input) and negative (7.5 μl of sterile water) amplification controls should be included with each amplification.
4. Prepare the Y-TPA amplification reaction mix with ~10 % excess to account for possible pipetting errors (volumes per sample): 12.5 μl Type-It Multiplex PCR master mix (1×), 2.5 μl Q-solution (0.5×), and 2.5 μl of Y-TPA primer mix. Vortex and briefly centrifuge.
5. Add 17.5 μl of Y-TPA reaction mix to each sample for a final reaction volume of 25 μl.
6. Amplify samples using the following cycling parameters: 95 °C 15 min; 15 cycles of 95 °C for 30 s, 60 °C for 90 s, 72 °C for 60 s; and 68 °C for 10 min. Amplified products should be stored at 4 °C.
7. Following amplification, purify Y-TPA amplification products using the MinElute PCR purification kit as described in Subheading 3.2 with the following modifications: (1) the entire 25 μl amplified product is purified which requires the additional of 125 μl of Buffer PB-I, (2) 25 μl of sterile water is used for elution (no concentration).

8. Label the appropriate number of 0.2 ml PCR tubes for the Yfiler® amplification (1/2 reaction volume, 12.5 μl).
9. Add 5 μl of purified Y-TPA product to the appropriate tubes. Positive (007 male DNA control DNA, 15 pg input) and negative (5 μl of sterile water) amplification controls should be included with each amplification.
10. Prepare the Yfiler® amplification mix (with ~10 % excess to account for possible pipeting errors) as follows (volumes per sample): 4.6 μl Yfiler® kit PCR reaction mix, 2.5 μl Yfiler® primer set, 0.4 μl AmpliTaq Gold DNA polymerase.
11. Add 7.5 μl of the prepared Yfiler® amplification mix to each sample.
12. Amplify samples using the following cycling conditions: 95 °C 11 min; 30 cycles of 94 °C for 1 min, 60 °C for 1 min, 72 °C for 1 min; and 60 °C for 80 min. Amplified products should be stored at 4 °C.
13. Proceed to Subheading 3.6.

3.6 Next Generation Y-STR Amplification (See Note 16)

3.6.1 AmpFlSTR® Yfiler® Plus PCR Amplification Kit

1. Thaw Yfiler® Plus Master Mix and Primer Set. Vortex and centrifuge briefly.
2. Label the appropriate number of 0.2 ml PCR tubes.
3. Add up to 1 ng of male DNA (maximum of 10 μl can be used for samples with undetectable or low male DNA quantitation values).
4. Add appropriate amount of low TE buffer to bring reaction volume to 10 μl. Positive (1 ng of 007 male control DNA, provided in kit) and negative (10 μl of low TE buffer) amplification controls should be included with each amplification.
5. Prepare the Yfiler® Plus amplification reaction mix with ~10 % excess to account for possible pipetting errors (volumes per sample): 10 μl Yfiler® Plus master mix and 5 μl Yfiler® Plus Primer Set. Vortex and briefly centrifuge.
6. Add 15 μl of Yfiler® Plus reaction mix to each sample for a final reaction volume of 25 μl.
7. Amplify samples using the following cycling parameters: 95 °C for 1 min; 30 cycles of 94 °C for 4 s, 61 °C for 1 min; and 60 °C for 22 min. Amplified products should be stored at 4 °C.
8. Proceed to Subheading 3.6.

3.6.2 PowerPlex® Y23 System (Promega)

1. Thaw the PowerPlex® Y23 5× Master Mix, PowerPlex® Y23 10× Primer Pair Mix, and Amplification Grade Water (provided in kit).
2. Label the appropriate number of 0.2 ml PCR tubes.

3. Add up to 0.5 ng of male DNA (maximum of 10 μl can be used for samples with undetectable or low male DNA quantitation values).
4. Add appropriate amount of amplification grade water to bring reaction volume to 17.5 μl. Positive (0.5 ng of 2800 M male control DNA, provided in kit) and negative (10 μl of amplification grade water) amplification controls should be included with each amplification.
5. Prepare the PowerPlex® Y23 reaction mix with ~10 % excess to account for possible pipeting errors (volumes per sample): 5 μl PowerPlex® Y23 5× Master Mix and 2.5 μl PowerPlex® Y23 10× Primer Pair Mix. Vortex and briefly centrifuge.
6. Add 7.5 μl of PowerPlex® Y23 reaction mix to each sample for a final reaction volume of 25 μl.
7. Amplify samples using the following cycling parameters: 96 °C 2 min; 30 cycles of 94 °C 10 s, 61 °C 1 min, 72 °C 30 s; 60 °C 20 min. Ensure that "max" mode is used as the ramp speed using a 9700 thermal cycler (Life Technologies). Amplified products should be stored at 4 °C.
8. Proceed to Subheading 3.6. Capillary Electrophoresis.

3.7 Capillary Electrophoresis

For Yfiler® or Yfiler® Plus, 1 μl of the amplified sample is added to 9.7 μl Hi-Di™ formamide and 0.3 μl of GeneScan™ 500 LIZ® (Yfiler®, G5 dye set) or 600 LIZ® (Yfiler® Plus, J6 dye set). For PowerPlex® Y23, 1 μl of the amplified sample is added to 10.0 μl Hi-Di™ formamide and 1.0 μl of CC5 ILS-500 (Any5Dye set). Samples were analyzed using an ABI Prism 3130 Genetic Analyzer (*see* **Notes 7** and **17**). Raw data are analyzed with the GeneMapper software (Life Technologies).

4 Notes

1. Numerous alternative DNA extraction methods and commerically available kits are available for DNA extraction/isolation. It may be possible to use these alternative extraction methods in place of organic DNA extractions. However, we have not evaluated any of the alternative approaches specifically for use with extended interval post-coital samples (≥6 days) and the subsequent Y-STR typing protocols described here. Therefore, the suitability of any alternative extraction method would need to be determined by the user. Additionally, for extended interval post-coital sample analysis, our standard extraction method is a manual organic DNA extraction with ethanol precipitation. The use of a differential extraction (separation of sperm and non-sperm cells) is not recommended for use with extended

interval post-coital samples (≥6 days). The protocols described here are intended for use with Y-STR typing strategies only as autosomal STR analysis on extended interval post-coital samples is unlikely to be successful and therefore separation of sperm and non-sperm cell fractions is not necessary.

2. The pH of the phenol-chloroform should not be below pH 6.7; otherwise DNA will be soluble in the organic phase and therefore not recoverable.
3. Quantification of male DNA can be performed prior to extract concentration and purification to determine if the sample contains sufficient amounts of male DNA for analysis prior to extract concentration and purification. This is not typical in extended interval post-coital samples ≥6 days. If desired, quantification can be performed before and after concentration.
4. In order to assess the amount of male and female DNA present in the sample, we recommend the use of a quantification system that permits a simultaneous quantification of both male and human DNA. Alternative DNA quantification kits are available that simultaneously assess male and human DNA [e.g., Quantifiler® Duo (Life technologies), Plexor® HY System (Promega), Investigator Quantiplex HYres kit (QIAGEN)]. We have not evaluated these alternative quantification kits. Therefore, the suitability of any alternative quantification method would need to be determined by the user.
5. Alternative real time PCR instruments are available and can be used in place of the 7500 Real Time PCR instrument indicated here. Users should always follow the manufacturer's recommended protocols for specific real time PCR instruments.
6. The use of the low TE buffer is recommended. We have also successfully used sterile 18.2 MΩ water in place of the low TE buffer. The low TE buffer can be prepared or purchased (e.g., Teknova). To prepare the low TE buffer: Add 10 ml of 1 M Tris–HCl pH 8.0, 0.2 ml 0.5 M EDTA pH 8.0, and 990 ml sterile water. Autoclave. Store at room temperature.
7. Capillary electrophoresis can be performed using alternative instruments (ABI Prism 310, 3130xl, 3500 Genetic analyzers) and performance-optimized polymers (POP-4, POP-6). Users should always refer to the manufacturer's recommended protocols for their instruments and the amplification kits. Electrophoretic conditions should be modified as needed by individual laboratories.
8. While some of the included protocols describe the use of commerically available kits, these protocols may contain modifications from the recommended manufacturer's instructions. Therefore, the user should always refer to the manufacturer's recommended protocol for further information and standard protocols.

9. For extended interval post-coital samples (≥6 days), we utilize an entire swab for extraction to maximize the amount of male DNA present in the sample. While the size of the sample used for analysis can be modified depending on the needs of individual laboratories, the maximum amount of the sample should be used in order to maximize the potential for successful DNA profiling results.
10. In order to maximize the amount of male DNA added to subsequent PCR reactions, we concentrate and purify our DNA extracts. This approach is used rather than reducing the volume of TE^{-4} used for re-solubilization in order to ensure efficient re-solubilization of the large amount of DNA present in non-differentially extracted samples (both male and female DNA present).
11. This protocol can be semi-automated using the QIACube (QIAGEN). Here we describe the full manual protocol.
12. Prior to extract concentration and purification, an aliquot of standard extract can be set aside for future use (or for comparison studies) if needed. Therefore the volume of DNA extraction used may vary. If the entire extract is to be concentrated and purified, 75 μl of DNA extract (1× volume) would be added to 375 μl (5× volume).
13. If the color of the sample and buffer PB-I changes to orange or violet, add 10 μl of 3 M sodium acetate pH 5.0 and mix.
14. Ensure that the pH of the water is between 7.0 and 8.5 for maximum elution efficiency. Alternatively, samples can be eluted into the provided Buffer EB or TE buffer (although the EDTA may inhibit subsequent enzymation reactions). The use of 13 μl of sterile water typically permits the recovery of ~12 μl (2 μl used for quantitation, 10 μl available for amplification reactions).
15. For extended interval post-coital samples (≥6 days), undetectable male DNA quantification values are often obtained. This does not preclude their use in the subsequent Y-TPA protocol. However, undetectable or extremely low male DNA quantitation values may not be successful with the next generation Y-STR amplification kits. For example, in our experience with the AmpFlSTR Yfiler® Plus amplification kit [29], we do not routinely obtain probative Y-STR profiles for samples with <0.002 ng/μl quantitation values (lowest quantitation standard value). We routinely test all samples regardless of the quantitation value, but the testing of samples with undetectable quantitation values is likely to be laboratory dependent.
16. Only standard or concentrated extracts should be amplified with the next generation Y-STR kits. The Y-TPA protocol is designed for use with Yfiler® or PowerPlex® Y. The Y-TPA

primer mix does not contain the new loci included in the next generation Y-STR kits.

17. Electrophoresis conditions for the 3130 Genetic Analyzer were as follows: 16 s injection time, 1.2 kV injection voltage, 15 kV run voltage, 60 °C, 20 min run time, dye set G5 (Yfiler®); 16 s injection time, 15 kV run voltage, 1.2 kV injection voltage, 60 °C, 25 min run time, J6 dye set (Yfiler® Plus); 5 s injection time, 3 kV injection voltage, 15 kV run voltage, 60 °C, 25 min run time, Any5Dye set (PowerPlex® Y23).

References

1. Hall A, Ballantyne J (2003) Novel Y-STR typing strategies reveal the genetic profile of the semen donor in extended interval post-coital cervicovaginal samples. Forensic Sci Int 136: 58–72
2. Mayntz-Press KA, Sims LM, Hall A et al (2008) Y-STR profiling in extended interval (> or =3 days) postcoital cervicovaginal samples. J Forensic Sci 53:342–348
3. Davies A, Wilson E (1974) The persistence of seminal constituents in the human vagina. Forensic Sci 3:45–55
4. Eungprabhanth V (1974) Finding of the spermatozoa in the vagina related to elapsed time of coitus. Z Rechtsmed 74:301–304
5. Morrison AI (1972) Persistence of spermatozoa in the vagina and cervix. Br J Vener Dis 48:141–143
6. Perloff WH, Steinberger E (1964) In vivo survival of spermatozoa in cervical mucus. Am J Obstet Gynecol 88:439–442
7. Rupp JC (1969) Sperm survival and prostatic acid phosphatase activity in victims of sexual assault. J Forensic Sci 14:177–183
8. Silverman EM, Silverman AG (1979) Presence of spermatozoa in cervicovaginal smears from young and old women. Exp Aging Res 5:155–159
9. Willott GM, Allard JE (1982) Spermatozoa--their persistence after sexual intercourse. Forensic Sci Int 19:135–154
10. Corach D, Filgueira RL, Marino M et al (2001) Routine Y-STR typing in forensic casework. Forensic Sci Int 118:131–135
11. Daniels DL, Hall AM, Ballantyne J (2004) SWGDAM developmental validation of a 19-locus Y-STR system for forensic casework. J Forensic Sci 49:668–683
12. Hall A, Ballantyne J (2003) The development of an 18-locus Y-STR system for forensic casework. Anal Bioanal Chem 376:1234–1246
13. Hanson EK, Ballantyne J (2004) A highly discriminating 21 locus Y-STR "megaplex" system designed to augment the minimal haplotype loci for forensic casework. J Forensic Sci 49:40–51
14. Hanson EK, Berdos PN, Ballantyne J (2006) Testing and evaluation of 43 "noncore" Y chromosome markers for forensic casework applications. J Forensic Sci 51:1298–1314
15. Hanson EK, Ballantyne J (2007) An ultra-high discrimination Y chromosome short tandem repeat multiplex DNA typing system. PLoS One 2:e688
16. Hanson EK, Ballantyne J (2010) The forensic application of Y-chromosome short-tandem repeats. In: Saferstein R (ed) Forensic science handbook volume III. Prentice Hall, Upper Saddle River, pp 436–466
17. Kayser M, Caglia A, Corach D et al (1997) Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110:125–129
18. Krenke BE, Viculis L, Richard ML et al (2005) Validation of a male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex. Forensic Sci Int 148:1–14
19. Mayntz-Press KA, Ballantyne J (2007) Performance characteristics of commercial Y-STR multiplex systems. J Forensic Sci 52: 1025–1034
20. Mulero JJ, Chang CW, Calandro LM et al (2006) Development and validation of the AmpFlSTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system. J Forensic Sci 51:64–75
21. Parson W, Niederstatter H, Brandstatter A et al (2003) Improved specificity of Y-STR typing in DNA mixture samples. Int J Legal Med 117:109–114
22. Prinz M, Sansone M (2001) Y chromosome-specific short tandem repeats in forensic casework. Croat Med J 42:288–291

23. Tsuji A, Ishiko A, Ikeda N et al (2001) Personal identification using Y-chromosomal short tandem repeats from bodily fluids mixed with semen. Am J Forensic Med Pathol 22: 288–291
24. Yoshida Y, Fujita Y, Kubo S (2004) Forensic casework of personal identification using a mixture of body fluids from more than one person by Y-STRs analysis. J Med Invest 51: 238–242
25. Ballantyne J (2013) DNA profiling of the semen donor in extended interval post-coital samples. NIJ Final Report 2007-DN-BX-K148. www.ncjrs.gov/pdffiles1/nij/grants/241299
26. Ballantyne J, van Daal A, Lubenow H (2013) Improved detection of male DNA in post-coital samples. NIJ Final Report 2007-DN-BX-K147. www.ncjrs.gov/pdffiles1/nij/grants/241298
27. Strom CM, Rechitsky S (1998) Use of nested PCR to identify charred human remains and minute amounts of blood. J Forensic Sci 43: 696–700
28. Ballantyne KN, Keerl V, Wollstein A et al (2012) A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. Forensic Sci Int Genet 6:208–218
29. Ballantyne J, Hanson E, Green R et al (2013) Enhancing the sexual assault workflow: testing of next generation DNA assessment and Y-STR systems. Forensic Sci Int Genet. doi:10.1016/j.fsigss.2013.10.117

Chapter 16

Analysis of Rapidly Mutating Y Chromosome Short Tandem Repeats (RM Y-STRs)

Sibte Hadi

Abstract

Rapidly mutating (RM) Y-STRs were recently identified and can help to differentiate between paternally related males in paternity and forensic casework. Normal Y-STRs are often used in casework due to their paternal inheritance, which can help to resolve kinship cases, and identify male components in male/female mixtures; however, more discriminating profiles are obtained if rapidly mutating Y-STRs are used. Previously two or three multiplex PCRs have been used to amplify 13 RM Y-STRs; here, an assay amplifying these 13 markers in a single multiplex PCR is described. Commercially available male control DNA samples have been genotyped during the validation of this assay, thus providing a tool for calibrating genotyping results. It is expected that the assay will provide a niche tool for genotyping casework samples.

Key words Y-STRs, Rapidly mutating Y-STRs, Multiplex PCR

1 Introduction

Y chromosome is the smallest chromosome in the human genome, with an average size of 60 Mb [1, 2]. Most of the Y chromosome consists of a non-recombinant region of the Y chromosome (NRY) [3]. The NRY is inherited intact through paternal lineages unless mutation/s have occurred. Because of such inheritance pattern, short tandem repeat (STR) markers located in the NRY region have become useful for applications including genetic structure studies, paternity testing, identification of disasters male victims, identification of male lineages for anthropology purposes, and the identification of male perpetrators in sexual assault criminal cases [4–10]. Y-STRs, which have an average mutation rate of about 10^{-3} per locus per generation [11, 12], have proven to be useful for forensic applications and have been included in several commercial Y-STR kits.

Although the greatest value of Y STRs is male specificity, this also turns into a major limitation due to the existence of

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_16,

an identical haplotype within a male lineage [13, 14]. This means that while currently used Y-STRs are able to reliably differentiate between different male lineages, they cannot resolve these lineages down to individual level in case of paternal relatives [15].

In order to enhance the discrimination capacity, Y-STR markers having a higher mutation rate were sought and in a study, mutation rates of 186 Y-STR loci were estimated in approximately 2000 father-son pairs from Poland and Germany [12]. As a result, 13 STR markers were discovered having mutation rates of 1×10^{-2} and higher. These comprised 12 tetranucleotide markers and 1 trinucleotide marker; 4 markers out of the 13 Y-STRs were multicopy markers and 9 were single copy markers. These 13 markers were termed rapidly mutating Y-STR (RM Y-STR) and their properties are summarized in Table 1 along with their mutation rates [12]. In a recent study, Ballantyne et al. [16] have employed three multiplex panels consist of 13 RM Y-STRs to investigate 604 unrelated male samples from 51 worldwide populations samples (HGDP-CEPH). As a result, 13 RM Y-STR loci showed higher haplotype diversity and haplotype discrimination capacity than that obtained using 17 Y-STRs available in Yfiler® multiplex kit (Life Technologies) [16]. In addition, this set of markers was extremely successful in differentiating between closely and distantly related males with 4.4-fold increase of average male relative differentiation relative to Yfiler® kit [16]. Moreover, in a worldwide-scale study of RM Y-STRs, more than 99 % of 12,272 investigated unrelated male individuals were completely individualized. Haplotype diversity in such large number of samples was found to be extremely high with haplotype sharing between 111 worldwide populations was almost absent, apart from 6 haplotypes which were shared between populations within the same geographical region [17]. Although the value of RM Y-STRs already been demonstrated, multiplexing the 13 RM YSTRs had not been reported until recently when a multiplex termed RM Yplex was developed that could be used in forensic casework alongside other human identification kits [18].

It should be noted that the interpretation of multi-copy RM YSTR markers DYF387S1, DYF399S1, DYF403S1, and DYF404S1 is often difficult for two reasons: firstly, the imbalance in the amplified peaks can affect the correct genotyping; and secondly, slippage peaks are common and these can be approximately 20 % of the allelic peaks. –1 stutter peaks in particular can be high and can pose difficulties in interpretation and need to be carefully assessed. +1 stutters are also present in these loci though these are rarely more than 5 % on the allelic peak. In mixed samples, where more than one male individual has contributed to the

Table 1
Properties of RM Y-STRs loci as adapted from Ballantyne et al. [16]

Locus	Repeat type	Repeat motif (variants in bold type)	Alleles ranges		Mutation rate
			Repeat number	Base pairs	
DYF387S1	Tetra, complex	$(AAAG)_3(GTAG)_1(GAAG)_4N_{16}(GAAG)_9$ $(\mathbf{AAAG})_{13}$	28–38	241–281	1.58×10^{-2}
DYF399S1	Tetra, complex	$(GAAA)_3\mathbf{N}_{7-8}(\mathbf{GAAA})_{10-23}$	10–23	261–313	7.73×10^{-2}
DYF403S1a	Tetra, complex	$(\mathbf{TTCT})_{10-17}\mathbf{N}_{2-3}(\mathbf{TTCT})_{3-17}$	12–39	310–438	3.10×10^{-2}
DYF403S1b	Tetra, complex	$(TTCT)_{12}N_2(\mathbf{TTCT})_8(\mathbf{TTCC})_9(\mathbf{TTCT})_{14}N_2(TTCT)_3$	40–59	414–490	$1.18\times10^{-}{}_2$
DYF404S1	Tetra, complex	$(\mathbf{TTTC})_{10-20}N_{42}(TTTC)_3$	10–20	171–211	1.24×10^{-2}
DYS449	Tetra, complex	$(\mathbf{TTTC})_{12-18}N_{22}(TTTC)_3N_{12}(\mathbf{TTTC})_{12-18}$	24–37	309–361	1.22×10^{-2}
DYS518	Tetra, complex	$(AAAG)_3(GAAG)_1(\mathbf{AAAG})_{14-22}(GGAG)_1(AAAG)_4N_6(\mathbf{AAAG})_{11-19}$	23–35	243–291	1.84×10^{-2}
DYS526a	Tetra, complex	$(CCCT)_3N_{20}(\mathbf{CTTT})_{11-17}(\mathbf{CCTT})_{6-10}N_{113}(\mathbf{CCTT})_{10-17}$	10–17	138–166	2.72×10^{-3}
DYS526b	Tetra, complex	$(CCCT)_3N_{20}(\mathbf{CTTT})_{11-17}(\mathbf{CCTT})_{6-10}N_{113}(\mathbf{CCTT})_{10-17}$	29–42	345–397	1.25×10^{-2}
DYS547	Tetra, complex	$(\mathbf{CCTT})_{9-13}T(\mathbf{CTTC})_{4-5}N_{56}(\mathbf{TTTC})_{10-22}N_{10}(CCTT)_4(TCTC)_1$ $(\mathbf{TTTC})_{9-16}N_{14}(TTTC)_3$	36–48	410–458	2.36×10^{-2}
DYS570	Tetra, simple	$(\mathbf{TTTC})_{14-24}$	10–21	246–286	1.24×10^{-2}
DYS576	Tetra, simple	$(\mathbf{AAAG})_{13-22}$	13–23	170–210	1.43×10^{-2}
DYS612	Tri, simple	$(CCT)_5(CTT)_1(TCT)_4(CCT)_1(\mathbf{TCT})_{19-31}$	14–31	187–255	1.45×10^{-2}
DYS626	Tetra, complex	$(\mathbf{GAAA})_{14-23}N_{24}(GAAA)_3N_6(GAAA)_5$ $(AAA)_1$ $(\mathbf{GAAA})_{2-3}(GAAG)_1(GAAA)_3$	11–23	221–269	1.22×10^{-2}
DYS627	Tetra, complex	$(AGAA)_3N_{16}(AGAG)_3(\mathbf{AAAG})_{12-24}N_{81}(AAGG)_3$	10–24	301–372	1.23×10^{-2}

sample, peak height imbalance will make assigning particular alleles to various contributors difficult [19]. However, an advantage of using multi-copy Y-STRs is that each copy has an independent mutation probability and hence increases the chances of mutation in accordance with the number of copies present in the marker. Therefore, these markers have high power of discrimination [19–21].

2 Materials

The method for the use of RM YPlex is described here for an optimal result in a forensic laboratory. We use reference material (saliva and blood) that is stored on FTA™ cards and describe the in-house process of extracting the DNA in addition to the PCR setup and analysis. We also present the haplotypes of three commercial male control DNA for calibration purposes.

2.1 FTA® Extraction: Reagents and Equipment

1. FTA™ Purification Reagent.
2. Harris Micro-Punch (1.2 mm).
3. Cutting Mat.
4. PCR tubes (200 µl).
5. TE buffer pH 8.0.
6. Heat block/incubator (56 °C).

2.2 PCR Amplification and Analysis

1. Fluorescently labeled primers (*see* Table 2).
2. Platinum® PCR Multiplex Master Mix (Life Technologies).
3. PCR Grade water.
4. PCR tubes (200 µl)—strips or plates can be used.
5. Hi-Di™ Formamide (Life Technologies).
6. POP-4™ or POP-6™ (either polymer can be used; with POP-6 the separation of peaks is better).
7. 50 cm capillary array.
8. Internal size standard GS600 LIZ™.
9. ABI 9700 or ABI Veriti (others have used ABI 2700) Thermocycler (*see* **Note 1**).
10. Capillary electrophoresis (CE) platform (we used ABI3500).
11. GeneMapper IDX or GeneMapper 3.2 Software for the analysis of CE data.
12. Control DNA: Taqman® (Life Technologies), 2800 M (Promega), 9948 (Promega).

Table 2
Forward and reverse primers of 13 RM Y-STR loci

Locus	Forward primer 5′–3′	Reverse primer 5′–3′	Fluorescent label
DYF387S1	ACAGAGCTAGATTCCATTTTACCC	GCCACAGTGTGAGAAGTGTGA	ATTO 565
DYF399S1	GGGTTTTCACCAGTTTGCAT	CCATGTTTTGGGACATTCCT	6-FAM
DYF403S1a/b	CAAAATTCATGTGGATAATGAG	ACAGAGCAGGATTCCATCTA	Yakima Yellow
DYF404S1	GGCTTAAGAAATTTCAACGCATA	CCATGATGGAACAATTGCAG	Yakima Yellow
DYS449	TGGAGTCTCTCAAGCCTGTTC	CCATTGCACTCTAGGTTGGAC	ATTO 565
DYS518	GGCAACACAAGTGAAACTGC	TCAGCTCTTACCATGGGTGAT	ATTO 550
DYS526a/b	TCTGGTGAACTGATCCAAACC	GGGTTACTTCGCCAGAAGGT	6-FAM
DYS547	TCCATGTTACTGCAAAATACAC	TGACAGAGCATAAACGTGTC	6-FAM
DYS570	CTGGCTGTGTCCTCCAAGTT	GGCAACCTAAGCTGAAATGC	ATTO 565
DYS576	GTTGGGCTGAGGAGTTCAATC	GGCAGTCTCATTTCCTGGAG	ATTO 550
DYS612	CCCCATGCCAGTAAGAATA	GTGAGGGAAGGCAAAAGAAAA	6-FAM
DYS626	GCAAGACCCCATAGCAAAAG	AAGAAGAATTTTGGGACATGTTT	Yakima Yellow
DYS627	GATGGGGAGGTTGCAGTAAG	TCTGTGAGTCCACTGGAGACC	ATTO 550

3 Methods

All normal precautions should be taken to minimize the potential for contamination, such as using dedicated pre- and post-PCR areas, equipment and reagents, wearing suitable personal protective equipment (PPE), and using controlled areas, such as laminar flow hoods/biosafety cabinets where appropriate.

3.1 Blood/Saliva Stained FTA Punch Purification (See Notes 2 and 3)

1. In a biosafety cabinet, remove the FTA™ card from packaging. Unfold and lay on clean cutting mat. Hold the card only on its edge, being sure to not touch the card directly.
2. Using a clean puncher for making 1.2 mm disc, press firmly with puncher on a clean FTA™ card twice to clean the puncher.
3. Using the clean puncher for making 1.2 mm disc, press firmly with puncher on area where the specimen is distributed on card.
4. Place FTA™ disc in a 0.2 ml PCR tube by clicking the end of puncher to release the disc. Clean puncher by punching the clean FTA™ card twice and repeat process (**steps 1**–**3**) until all specimens have been punched.
5. Finally, make a final card punch with a cleaned puncher in the area that is fairly outside the specimen area. Place this disc into a PCR tube labeled as a negative extraction control.
6. Add 200 μl of FTA™ purification reagent to each PCR tube.
7. Incubate at room temperature for 15 min, then remove and discard spent FTA™ reagent.
8. Add 200 μl TE buffer pH 8.0 to each PCR tube.
9. Incubate for 5 min at room temperature, then remove and discard TE buffer.
10. Repeat **steps 8** and **9** once for a total of two washes with TE buffer.
11. Dry samples at 56 °C in heat block or incubation chamber with caps open until the discs are dry (approximately 30 min).
12. Set up PCR as below adding one FTA disc (two discs can be added to the 15 μl reaction volume if the DNA concentration is low).

3.2 PCR Amplification

1. Prepare PCR primers at a working concentration of 20 μM in sterile, UV treated 2 ml screw cap vials. From these stocks, prepare a PCR primer mix to achieve the concentrations of various primer pairs as per Table 3.
2. Use 3 μl of this primer mix for each sample when preparing a PCR master mix (*see* **Note 4**).
3. Make a master mix as follows to a reaction volume of 15 μl as follows:

   7 μl of Platinum® PCR Multiplex Master Mix (Life Technologies) (*See* **Note 5**)

Table 3
Working primer concentrations in a 15 μl reaction volume; volumes of forward (F) and reverse (R) primers required in a master mix are given for a 20 μM primer stock (*see* Note 4)

Marker	F primer (μM)	R primer (μM)	F primer (μl)	R primer (μl)
DYF387S1	0.05	0.05	0.038	0.038
DYF399S1	0.06	0.06	0.045	0.045
DYF403S1a/b	0.40	0.40	0.300	0.300
DYF404S1	0.11	0.11	0.083	0.083
DYS449	0.06	0.06	0.045	0.045
DYS518	0.12	0.12	0.090	0.090
DYS526a/b	0.40	0.40	0.300	0.300
DYS547	0.40	0.40	0.300	0.300
DYS570	0.04	0.04	0.030	0.030
DYS576	0.03	0.03	0.023	0.023
DYS612	0.10	0.10	0.075	0.075
DYS626	0.07	0.07	0.053	0.053
DYS627	0.20	0.20	0.150	0.150

3.0 μl of the 13 RM Y-STRs primer mix (Table 3)
1–5 μl of DNA template (depending on concentration) (*see* **Note 6**).
1–4.0 μl PCR grade water to a final volume of 15 μl.

In casework there can be situations when a large volume of extract needs to be used as the template concentration is low. This assay allows a maximum volume of 6.5 μl of DNA template (when using 40 μM primer concentration) to be added to the assay which is comparable to commercially available assays (*see* **Notes 4**, **6**, and **7**).

4. Prepare the reactions taking usual anticontamination measures. Amplify the samples in a thermocycler using GeneAmp 9700 or Veriti (Life Technologies) at an initial denaturation temperature of 95 °C for 10 min, followed by 12 cycles at 94 °C for 30 s, 58 °C for 45 s, 72 °C for 60 s and then 20 cycles at 94 °C for 30 s, 55 °C for 45 s, 72 °C for 60 s. Use a final extension temperature of 72 °C for 45 min for complete adenylation.

3.3 Capillary Electrophoresis and Analysis

1. Prepare a master mix using 9.6 μl Hi-Di™ Formamide (Life Technologies), 0.4 μl GS600 LIZ™ size standard for each sample adding for two extra samples.

2. Aliquot 10 μl in each well if using a plate or each tube if using tubes for loading. Add 1 μl PCR product to each well. Add a ladder in each injection (*see* **Note 8**).
3. The prepared samples should be injected into the capillary/ies of the sequencer for 10 s at 3 kV and electrophoresis performed at 15 kV for 44 min, at a run temperature of 60 °C using the POP-6™ sieving polymer (Life Technologies) (*see* **Note 9**).
4. Import the samples in a new project in Genemapper 3.2 or IDX versions. The analysis parameters should be kept at default except that the minimum peak height should be kept at 50 RFU. For analyzing the samples use the G5 matrix standard (*see* **Note 10**).
5. Call alleles using the conservative method of calling an allele without making use of relative peak height in the case of multicopy markers (*see* **Note 11**).

3.4 Genotyping of Commercial Controls

Three commercially available controls 9948, Taqman Male Control, and 2800 have been genotyped for the purposes of calibration of results for different laboratories a number of times in different laboratories by different analysts using different PCR machines (Veriti and ABI 9700) (Table 4; Fig. 1). A minimum of two of these controls is recommended to be used in every amplification batch for assisting in the allele calls.

Table 4
Genotype data for three commercially available male DNA controls. 9948 and 2800 controls are available from Promega Corporation; Taqman® control is available from Life Technologies

Locus	9948	TaqMan®	2800 M
DYF387S1	35:38	36:39	37:38
DYF399S1	22:25.1	19:23	24:25.1:26.1
DYF403S1a	10:15:16	11.2:12.2:15	11:14:17
DYF403S1b	49	54	46
DYF404S1	12:14	15:18	13:16
DYS449	31	30	35
DYS518	38	38	36
DYS526a	14	13	12
DYS526b	36	34	36
DYS547	48	48	45
DYS570	18	19	17
DYS576	16	15	18
DYS612	37	36	35
DYS626	28	29	29
DYS627	22	21	22

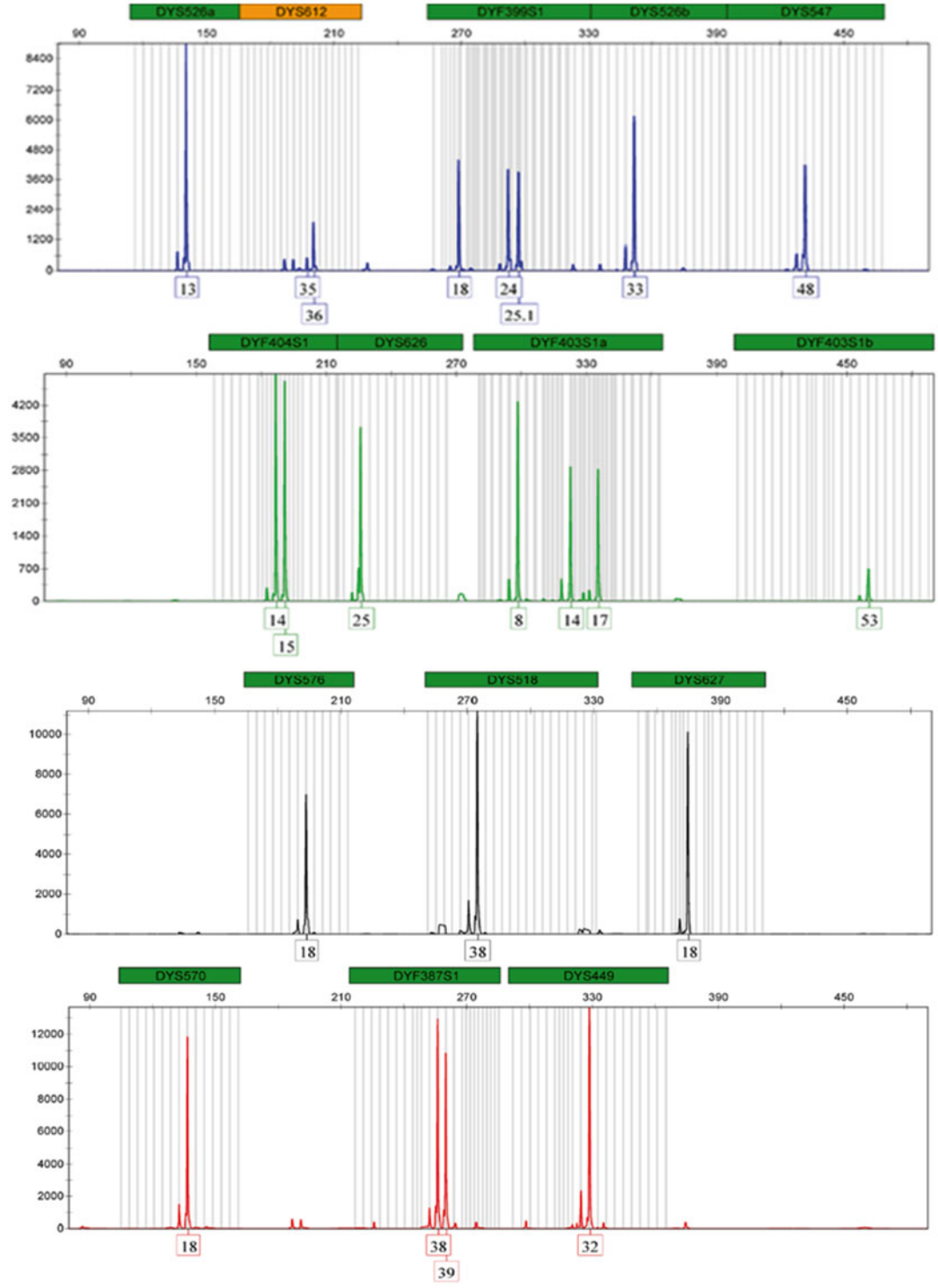


Fig. 1 An electropherogram showing male profile generated using RM YPlex for 13 rapidly mutating Y-STRs. Sample was injected in the 8-Capillary 50 cm array for 10 s at 3 kV. Separation was performed at 15 kV for 44 min at a run temperature of 60 °C using the POP-6™ polymer. The sample was analyzed using G5 matrix

4 Notes

1. We have validated RM YPlex using ABI 9700 and Veriti Thermal cyclers. Other PCR machines can be used though it would need a verification to be done by individual laboratories.
2. If there is a need to extract DNA from the FTA card disc a standard Phenol Chloroform method can be used (*see* Chapter 5).
3. We use the Qiagen mini kit for extraction of DNA using standard protocols. Any procedure that provides DNA that is suitable for PCR analysis can be used.
4. For using larger volumes of DNA extracts, prepare a 40 μM working primer concentration. This will reduce the volume of each primer to half and the total primer mix volume will be reduced to 1.5 μl. This way up to 6.5 μl of DNA extract could be added to the PCR. Keep the PCR volume at 15 μl.
5. RM YPlex has been optimized using other enzyme mixes like Qiagen PCR mix and Phusion Flash allowing laboratories to use PCR mixes other than Platinum PCR mix. The data generated using other PCR mixes needs careful calibration using the controls indicated in Table 4.
6. The optimum amount of template DNA is 500 pg: we routinely obtain full profiles down to 62.5 pg.
7. Test the assay and adjust the primer concentrations as per requirement keeping the final volume same. The total volume of the primer mix can be increased up to 1.73 μl when using primers at 40 μM concentration.
8. Currently a sequenced ladder is not available for general use by all laboratories. We can supply the panels and bin sets for the 13 RM Loci but would recommend preparing a ladder for usage when using the RM Yplex. This can be done by profiling approximately 100 individuals to gather the data for commonly occurring alleles in the male population, preparing a mixture of alleles and running it as a ladder for individual loci. The commercial male controls can help in calibrating the system in individual laboratories.
9. RM Yplex has been validated using POP-6™ polymer on ABI 3500. If using POP-4™ run time on ABI 3500 will need to be adjusted by individual laboratories when performing an internal validation.
10. Use of G5 matrix leads to some pull up peaks in the blue panel so exercise care while interpreting the data. Minimum RFUs have to be assessed in each laboratory.
11. In some loci split peaks can be a problem if using POP-6. Using POP-4™ can resolve such issues. Since most forensic laborato-

ries use POP-4™ this phenomenon may not be exhibited. Interpretation of data needs some experience in using the multiplex assay and confidence in calling the peaks of the multi-copy markers. In case of doubt re-amplify and re-analyze the sample.

References

1. Buehler EM (1980) A synopsis of the human Y chromosome. Hum Genet 55:145–175
2. Skaletsky H, Kuroda-Kawaguchl T, Minx PJ, Cordum HS, Hlllier L, Brown LG et al (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–837
3. Quintana-Murci L, Krausz C, McElreavey K (2001) The human Y chromosome: function, evolution and disease. Forensic Sci Int 118:169–181
4. Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G et al (1997) Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110:125–133
5. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH et al (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361
6. Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M (2001) Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat Genet 29:20–21
7. Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H et al (2001) Hierarchical patterns of global human Y-chromosome diversity. Mol Biol Evol 18:1189–1203
8. Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612
9. Roewer L, Croucher PJP, Willuweit S, Lu TT, Kayser M, Lessig R et al (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. Hum Genet 116:279–291
10. Shi W, Ayub Q, Vermeulen M, Shao RG, Zuniga S, Van Der Gaag K et al (2010) A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. Mol Biol Evol 27:385–393
11. Goedbloed M, Vermeulen M, Fang RN, Lembring M, Wollstein A, Ballantyne K et al (2009) Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR® Yfiler® PCR amplification kit. Int J Legal Med 123:471–482
12. Ballantyne KN, Goedbloed M, Fang R, Schaap O, Lao O, Wollstein A et al (2010) Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. Am J Hum Genet 87:341–353
13. Mulero JJ, Chang CW, Calandro LM, Green RL, Li Y, Johnson CL et al (2006) Development and validation of the AmpFlSTR® Yfiler™ PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system. J Forensic Sci 51:64–75
14. Thompson R, Zoppis S, McCord B (2012) An overview of DNA typing methods for human identification: past, present, and future. Methods Mol Biol 830:3–16
15. Gusmão L, Sánchez-Diz P, Alves C, Beleza S, Lopes A, Carracedo A et al (2003) Grouping of Y-STR haplotypes discloses European geographic clines. Forensic Sci Int 134: 172–179
16. Ballantyne KN, Keerl V, Wollstein A, Choi Y, Zuniga SB, Ralf A et al (2012) A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. Forensic Sci Int Genet 6:208–218
17. Ballantyne KN, Ralf A, Aboukhalid R, Achakzai NM, Anjos MJ, Ayub Q et al (2014) Toward male individualization with rapidly mutating Y-chromosomal short tandem repeats. Hum Mutat 35:1021–1032
18. Alghafri R, Goodwin W, Ralf A, Kayser M, Hadi S (2015) A novel multiplex assay for simultaneously analysing 13 rapidly mutating Y-STRs. Forensic Sci Int Genet 17:91–98
19. Butler JM, Schoske R (2005) U.S. population data for the multi-copy Y-STR locus DYS464. J Forensic Sci 50:975–977
20. Berger B, Niederstätter H, Brandstätter A, Parson W (2003) Molecular characterization and Austrian Caucasian population data of the multi-copy Y-chromosomal STR DYS464. Forensic Sci Int 137:221–230
21. Jacobs M, Janssen L, Vanderheyden N, Bekaert B, Van de Voorde W, Decorte R (2009) Development and evaluation of multiplex Y-STR assays for application in molecular genealogy. Forensic Sci Int Genet 2:57–59

Chapter 17

A Practical Guide to the HIrisPlex System: Simultaneous Prediction of Eye and Hair Color from DNA

Susan Walsh and Manfred Kayser

Abstract

The HIrisPlex system, which consists of two parts, allows the simultaneous prediction of eye and hair color from DNA, e.g., samples extracted from crime scene evidence. The first part is a highly sensitive multiplex genotyping assay consisting of 24 DNA markers using SNaPshot™ chemistry, for analysis on all Capillary Electrophoresis machines. The second part consists of statistical models that respectively establish eye and hair color prediction probabilities from complete and incomplete genotype profiles using parameters generated from large genotype and phenotype databases. This combined prediction tool constitutes the online system freely available to users. Here we provide a practical guide on how to use the HIrisPlex system for forensic and other DNA applications.

Key words HIrisPlex, Eye color, Hair color, Prediction, SNP, Forensic DNA Phenotyping

1 Introduction

DNA phenotyping refers to the prediction of a phenotype from a genotype. When used in the forensic context, i.e., Forensic DNA Phenotyping (FDP), externally visible phenotypes are predicted. FDP tools are used for investigative purposes to find unknown sample donors in cases where standard STR profiling from DNA does not lead to conclusive identification [1–3]. In and out of the forensic field, these tools are used to assign appearance traits that are unavailable from skeletal remains to deceased persons for anthropological [4, 5] or investigative use [6]. Furthermore, DNA phenotyping involving diseases as phenotypes, i.e., the prediction of disease risks from genotypes, is also relevant in the emerging field of personalized medicine [7, 8]. DNA phenotyping has gained a lot of interest and potential within the last few years due to success in the association of DNA variants with particular human

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_17,

phenotypes via genome-wide or candidate gene studies, such as eye color [9–18] and hair color [9–11, 15, 17, 19–23].

In the case of eye color, combining these scientific discoveries and assessing their predictive capabilities [24] led to the development of several predictive tools such as IrisPlex [25], amongst others [26, 27]. Combining eye color with hair color DNA prediction [28] then led to the formation of the HIrisPlex system [29]. The HIrisPlex system, capable of simultaneously predicting eye and hair color from DNA, consists of a single multiplex genotyping assay for 24 eye and/or hair color informative DNA variants, and statistical prediction models, one for eye and one for hair color that are based on large underlying genotype and phenotype databases. The HIrisPlex system has undergone strict forensic developmental validation guidelines to ensure it is one of the most reliable genotyping assays currently available for eye and hair color marker genotyping [30]. What adds to this tools efficiency and ease of use is the freely available online interactive prediction tool found at http://www.erasmusmc.nl/fmb/resources/Irisplex_HIrisPlex/

that is capable of predicting eye and hair color from the full 24-marker profile. This online tool is also capable of producing a conservative prediction with partial profiles (profiles with 23 variants or less) while additionally revealing its loss in prediction accuracy due to the missing loci.

Accuracy estimates using a previous model for eye color can be found in Walsh et al. 2012 [31] and for both eye and hair color in Walsh et al. 2013 [29]. However, there is now an enhanced model [30] that currently uses the largest known database of eye (9188 individuals) and hair color (1601 individuals) phenotype and genotype information with Area Under the receiver operating Curve (AUC) measurements of 0.94 for blue, 0.74 for intermediate, 0.95 for brown eye color, and 0.81 for blond, 0.75 for brown, 0.92 for red, 0.85 for black hair color, and 0.90 for hair color shade. The AUC statistic describes on average (across all individuals within the dataset) prevalence-adjusted prediction accuracy measures that run on a scale from 0.5, meaning random phenotype prediction, i.e., no genotype information is relevant, to 1.0, meaning completely accurate phenotype prediction, i.e., every time one predicts, the correct phenotype is obtained. The HIrisPlex systems current evaluation revealed that concurrent eye and hair color can be accurately and conservatively predicted in three out of four cases. Based on a small 119 test set, precision accuracy for eye color is 84 % (93 % for blue and brown), and for hair color is on average 73 %.

The presently achieved eye and hair color prediction accuracies shall improve in the coming years as the genetic basis of eye and hair color variation is understood more completely. Additional DNA markers found to independently contribute towards eye and hair color prediction, usually identified through genetic association studies, shall add greatly to such prediction systems together with epistatic

effects [32, 33]. Prediction improvements are particularly needed for the non-blue and non-brown eye colors, where current systems including Iris/HIrisPlex are not performing at an optimum level. Some developments are currently going on in this direction [26, 34, 35], but much more basic research is needed. Furthermore, age-related changes in hair color such as blond in childhood to brown/black in adulthood need to be understood on the molecular biology level, and at best including these identified biomarkers in the prediction systems to improve hair color hair color prediction in individuals that have undergone such a change. Notably, the HIrisPlex (and other currently available systems) predict hair color in early life, while hair color change with age is not predictable with these systems.

Here we describe what is needed and how to run the HIrisPlex system, from SNP genotyping using the HIrisPlex assay to inserting genotyped alleles into the prediction model and interpreting the prediction results. This tool has been successfully used for the prediction of eye and hair color from DNA in many cases, from aDNA to modern-day casework [6], so we anticipate that a full breakdown of the tools usage from start to finish will be of great benefit to many users. We also offer several tips and notes to try to ensure that the HIrisPlex system functions at its most optimum level for each user.

2 Materials

2.1 HIrisPlex Multiplex PCR Assay

1. HIrisPlex PCR primers (*see* Table 1).
2. 1× TE buffer solution: 10 mM Tris–HCl, 1 mM ethylenediamine tetra-acetic acid (EDTA), pH 8.
3. AmpliTaq Gold® DNA polymerase containing 10× PCR buffer and 25 mM $MgCl_2$ (Life Technologies).
4. PCR nucleotide mix, 10 mM (Invitrogen, Groningen, The Netherlands).
5. ddH_2O.
6. Bovine Serum Albumen (BSA) (10 mg/ml).
7. 0.2 ml tubes (single).
8. PCR instrument GeneAmp 9600 thermal cycler (Life Technologies).

2.2 HIrisPlex Multiplex Primer Extension Assay

1. HIrisPlex High-performance liquid chromatography-purified SBE primers (*see* Table 1).
2. SNaPshot ® Multiplex Chemistry (Life Technologies).

2.3 Purification of Amplification Products (PCR and SBE)

1. Exonuclease Shrimp Alkaline Phosphatase—ExoSAP-IT (USB Corporation, Cleveland, OH).
2. Shrimp Alkaline Phosphatase (SAP) (USB Corp.).

Table 1
Information on the 24 DNA variants that make up the HIrisPlex developmentally validated assay (Walsh 2014), including PCR and single base extension (SBE) primer sequences

Assay position	SNP	PCR primers			Product size (bp)	SBE primers	Product size (bp)
1	**N29insA**	MC1Rset1F	**Set1**	GCAGGGATCCCAGAGAAGAC	117	CCAGCTGGGGCTGGCTGCCAA	**21**
2	**rs11547464**	MC1Rset1R		TCAGAGATGGACACCTCCAG		ttttttttttttGCCATCGCCGTGGACC	**28**
3	**rs885479**	MC1Rset2F	**Set2**	CTGGTGAGCTTGGTGGAGA	158	ttttttttttttttttttGATGGCCGCAACGGCT	**34**
4	**rs1805008**	MC1Rset2R		TCCAGCAGGAGGATGACG		ttttttttttttACAGCATCGTGACCCTGCCG	**33**
5	**rs1805005**	MC1Rset3F	**Set3**	GTCCAGCCTCTGCTTCCTG	147	ttttttttttttttTGGTGGAGAACGCGCTGGTG	**35**
6	**rs1805006**	MC1Rset3R		AGCGTGCTGAAGACGACAC		ttttttttttttttttttttCTGCCTGGCCTTGTCGGA	**38**
7	**rs1805007**	MC1Rset4F	**Set4**	CAAGAACTTCAACCTCTTTCTCG	106	ttttttttttttttttttttttttttCTCCATCTTCTACGCACTG	**44**
8	**rs1805009**	MC1Rset4R		CACCTCCTTGAGCGTCCTG		ttttttttttttttttttttttttttttttttATCTGCAATGCCATCATC	**48**
9	**Y152OCH**					ttttttttttttttttttttttttttttttttCATCTTCTACGCACTGCGCTA	**51**
10	**rs2228479**					ttttttttttttttttttttttttttttttttttttttCTGGTGAGCGGGAGCAAC	**54**
11	**rs1110400**					ttttttttttttttttttttttttttttttttttt CTTCTACGCACTGCGCTACCACAGCA	**56**
12	**rs28777**	rs28777_F	**Set5**	TACTCGTGTGGGAGTTCCAT	150	ttttttttttttttttttttttttttttttttttttttttttCATGTG ATCCTCACAGCAG	**58**
		rs28777_R		TCTTTGATGTCCCCTTCGAT			
13	**rs16891982**	Rs16891982_F	**Set6**	TCCAAGTTGTGCTAGACCAGA	128	ttttttttttttttttttttttttttttttttttttttttttttttttttAAACACG GAGTTGATGCA	**62**
		Rs16891982_R		CGAAAGAGGAGTCGAGGTTG			
14	**rs12821256**	rs12821256_F	**Set7**	ATGCCCAAAGGATAAGGAAT	118	ttttttttttttttttttttttttttttttttttttttttGGAGCCAAGGGCATGT TACTACGGCAC	**66**
		rs12821256_R		GGAGCCAAGGGCATGTTACT			

continued

15	**rs4959270**	Rs4959270_F	**Set8**	TGAGAAATCTACCCCCACGA	140	ttttttttttttttttttttttttttttttttttttttttGGAACACATCCAAACT ATGACACTATG	**68**
		Rs4959270_R		GTGTTCTTACCCCCTGTGGA			
16	**rs12203592**	rs12203592_F	**Set9**	AGGGCAGCTGATCTCTTCAG	126	ttttttttttttttttttttttttttttttttttt ttttttttttTCCACTTTGGTGGGTAAAAGAAGG	**69**
		rs12203592_R		GCTTCGTCATATGGCTAAACCT			
17	**rs1042602**	rs1042602 _F	**Set10**	CAACACCCATGTTTAACGACA	124	ttttttttttttttttttttttttttttttttt ttttttttttttttttttTCAATGTCTCTCCAGATTTCA	**73**
		rs1042602 _R		GCTTCATGGGCAAAATCAAT			
18	**rs1800407**	rs1800407_F	**Set11**	AAGGCTGCCTCTGTTCTACG	124	ttttttttttttttttttttttttttttttttt ttttttttttttttttttttttttttGCATACCGGCTCTCCC	**77**
		rs1800407_R		CGATGAGACAGAGCATGATGA			
19	**rs2402130**	rs2402130_F	**Set12**	ACCTGTCTCACAGTGCTGCT	150	ttttttttttttttttttttttttttttttt ttttttttttttttttttttttttTGAACCATACGGAGCCCGTG	**80**
		rs2402130_R		TTCACCTCGATGACGATGAT			
20	**rs12913832**	rs12913832_F	**Set13**	TCAACATCAGGGTAAAAATCATGT	150	ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt- ttttTAGCGTGCAGAACTTGACA	**83**
		rs12913832_R		GGCCCCTGATGATGATAGC			
21	**rs2378249**	rs2378249_F	**Set14**	CGCATAACCCATCCCTCTAA	136	ttttttttttttttttttttttttttttttttttttttttttttttttt- ttttttttttttttttCCACACCTCTCCTCAGCCCA	**86**
		rs2378249_R		CATTGCTTTTCAGCCCACAC			
22	**rs12896399**	Rs12896399_F	**Set15**	CTGGCGATCCAATTCTTTGT	125	ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt ttttTCTTTAGGTCAGTATATTTTGGG	**88**
		Rs12896399_R		GACCCTGTGTGAGACCCAGT			
23	**rs1393350**	Rs1393350_F	**Set16**	TTTCTTTATCCCCCTGATGC	124	ttttttttttttttttttttttttttttttttttttttttttttttttttttttt ttttttttttttCATTTGTAAAAGACCACACAGATTT	**92**
		Rs1393350_R		GGGAAGGTGAATGATAACACG			
24	**rs683**	rs683_F	**Set17**	CACAAAACCACCTGGTTGAA	138	ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttGCTT TGAAAAGTATGCCTAGAACTTTAAT	**95**
		rs683_R		TGAAAGGGTCTTCCCAGCTT			

2.4 Electrophoresis and Genotyping Software

1. HiDi formamide (Invitrogen).
2. GeneScan™ Liz 120™ dye Size Standard (Life Technologies).
3. Capillary Electrophoresis platform for detection of fluorescently labeled oligonucleotides with preloaded DS-02/Dye set E5 matrix (Life Technologies).
4. Computer with Genotyping software, GeneMapper™ (Life Technologies) or GeneMarker™ (Softgenetics, State College, PA).

2.5 HIrisPlex Online Prediction Tool

1. Computer with Internet access to site http://www.erasmusmc.nl/fmb/resources/Irisplex_HIrisPlex/.

3 Genotyping Methods

3.1 HIrisPlex Assay

The HIrisPlex assay [29] that has been forensically validated [30] enables users to simultaneously predict the eye and hair color of an individual from DNA using extremely low input amounts (down to 64 pg DNA input). In our experience, to achieve optimum results whilst using the HIrisPlex assay, it is essential that newly prepared primers and reagents are used. Therefore adequate preparation of quantities needed for a batch run and freezing of aliquots is recommended. dNTPs in particular are sensitive to repeated freeze-thaw, so it is essential to make up all working reagents in small tubes and only use the tube once. To begin using the HIrisPlex assay, the following primers (Table 1) must be ordered following our recommended quality control assessments that are done by the primer provider/company (i.e., HPLC purified (*see* **Note 1**)). Input DNA should be >64 pg DNA (*see* **Note 2**), and of reasonable quality; DNA extract from single source (*see* **Note 3**) and of sufficient fragment length (>150 bp), i.e., not severely degraded (*see* **Note 4**).

3.2 Primer Preparation

Make sure all tubes are labeled with its primer name and the date it was opened when making up primer stocks and working solutions. If you order primers lyophilized, then make sure you add the required amount of 1× TE Buffer or water to the tubes (*see* **Note 1**). Ensure the solution is thoroughly mixed by vortexing and centrifuging to remove liquid from the tops of the tubes. Make aliquots of these stocks to ensure your primers are not affected by subsequent freeze/thaw and store in the freezer. When making up working primer solutions, ensure you make enough to cover the runs needed for your current workload and store in the fridge at 4 °C. Working solutions should not be kept in the fridge for longer than 10 days or else PCR reactions may be affected. Primers should be prepared according to Table 2 for easy dispensing (*see* **Note 5**); an example of a 10 individual mix of primer quantities is also provided in the table.

Table 2
Information on PCR and SBE working and reaction concentrations/volumes used in the HIrisPlex assay (Walsh 2014)

No.	SNP	Primer set	PCR primer working solution (μm)	Reaction concentration (μm)	Forward Vol(μl)	Reverse Vol. (μl)	Total vol (μl)	10 reactions	SBE Dir.	SNPs	Reaction concentration (μm)	of 10 μm(μl) solution	of 20 μm(μl) solution	of 50 μm (μl) solution	10 reactions
1	N29insA	**MC1Rset1** (1 F&R)	50	0.5	0.1	0.1	0.2	2	1 F	C/A	1.3			0.13	1.3
2	rs11547464	**MC1Rset3** (3 F&R)	50	0.5	0.1	0.1	0.2	2	2 F	G/A	0.1	0.05			0.5
3	rs885479	MC1Rset3							3R	C/T	1.25			0.125	1.25
4	rs1805008	MC1Rset3							4 F	C/T	0.375			0.0375	0.375
5	rs1805005	**MC1Rset2** (2 F&R)	50	0.5	0.1	0.1	0.2	2	5 F	G/T	1			0.1	1
6	rs1805006	MC1Rset2							6 F	C/A	0.8			0.08	0.8
7	rs1805007	MC1Rset3							7 F	C/T	1.1			0.11	1.1
8	rs1805009	**MC1Rset4** (4 F&R)	50	0.4	0.08	0.08	0.16	1.6	8 F	G/C	0.12		0.03		0.3
9	Y152OCH	MC1Rset3							9 F	C/A	0.5			0.05	0.5
10	rs2228479	MC1Rset2							10 F	G/A	0.375			0.0375	0.375
11	rs1110400	MC1Rset3							11 F	T/C	0.1			0.01	0.1
12	rs28777	rs28777 - 5 F+R	50	0.4	0.08	0.08	0.16	1.6	12 F	A/C	1.2			0.12	1.2
13	Rs16891982	Rs16891982 - 6 F+R	50	0.4	0.08	0.08	0.16	1.6	13 F	G/C	0.9			0.09	0.9
14	rs12821256	rs12821256 - 7 F+R	50	0.4	0.08	0.08	0.16	1.6	14R	A/G	0.12		0.03		0.3

(continued)

No.	SNP	Primer set	PCR primer working solution (μm)	Reaction concentration (μm)	Forward Vol(μl)	Reverse Vol. (μl)	Total vol (μl)	10 reactions	SBE Dir.	SNPs	Reaction concentration (μm)	of 10 μm(μl) solution	of 20 μm(μl) solution	of 50 μm (μl) solution	10 reactions
15	rs4959270	rs4959270 - 8 F+R	50	0.4	0.08	0.08	0.16	1.6	15 F	C/A	0.3			0.03	0.3
16	rs12203592	rs12203592 - 9 F+R	50	0.4	0.08	0.08	0.16	1.6	16 F	C/T	0.2			0.02	0.2
17	rs1042602	rs1042602 - 10 F+R	50	0.4	0.08	0.08	0.16	1.6	17R	G/T	1.25			0.125	1.25
18	rs1800407	rs1800407 - 11 F+R	50	0.4	0.08	0.08	0.16	1.6	18 F	G/A	0.1			0.01	0.1
19	rs2402130	rs2402130 - 12 F+R	50	0.4	0.08	0.08	0.16	1.6	19 F	A/G	0.75			0.075	0.75
20	rs12913832	rs12913832 - 13 F+R	50	0.4	0.08	0.08	0.16	1.6	20R	C/T	1			0.1	1
21	rs2378249	rs2378249 - 14 F+R	50	0.4	0.08	0.08	0.16	1.6	21R	T/C	0.1			0.01	0.1
22	Rs12896399	Rs12896399 - 15 F+R	50	0.4	0.08	0.08	0.16	1.6	22 F	G/T	1			0.1	1
23	Rs1393350	Rs1393350 - 16 F+R	50	0.4	0.08	0.08	0.16	1.6	23R	C/T	1.1			0.11	1.1
24	rs683	rs683 - 17 F+R	50	0.4	0.08	0.08	0.16	1.6	24R	T/G	0.3			0.03	0.3
					Primer vol. 1st PCR	**Total**	**2.84**	**28.4**				**Primer vol. SBE**	**Total**	**1.61**	**16.1**

3.3 PCR Multiplex Reaction Mix

1. Prepare a MasterMix1 excel sheet of the total number of samples required for PCR amplification. Include volume for three additional samples. These are for additional pipetting volume, one negative control (ddH_2O) and one positive control, i.e., 9947A (*see* **Note 6**).
2. Follow the sheet and make a MasterMix1 containing:

   2.84 μL total of PCR Primers as per Table 2.

   1 μL of 10× PCR Gold Buffer (No Mg).

   1 μL of $MgCl_2$ (25 mM).

   0.22 μL dNTPs (10 mM each).

   0.3 μL Taq Gold Polymerase (5U/μL).

   0.7 μL BSA (10 mg/ml).

   2.94 μL H_2O.

   per each reaction required.
3. Add 9 μL of MasterMix1 to each 1 μL DNA sample in a 0.2 ml PCR tube.
4. Vortex and spin down.

3.4 Multiplex PCR Conditions

1. Program the thermocycler for the following conditions:

   95 °C for 10 min.

   33 cycles of 95 °C for 30 s and 61 °C for 30 s.

   5 min at 61 °C.

   15 °C forever.

3.5 Purification of the PCR Product

1. Transfer 5 μL PCR product to fresh 0.2 ml tubes (*see* **Note 7**).
2. Add 2 μL Exosap (*see* **Note 7**) to the PCR product (per sample).
3. Vortex and pulse spin down.
4. Incubate in PCR machine.

   37 °C for 45 min.

   80 °C for 15 min.

   15 °C forever.

3.6 Single Base Extension (SBE) Reaction

1. Prepare a MasterMix2 excel sheet of the total number of samples required for SBE amplification. Include volume for three additional samples. These are for additional pipetting volume, one negative control and one positive control (*see* **Note 8**).
2. Follow the sheet and make a MasterMix2 containing;

   (a) 1.61 μL total of SBE Primers as per Table 2.

   (b) 1 μL SNaPshot Reaction Mix.

(c) 0.39 μL H_2O.

(d) per each reaction required.

- Add 3 μL of MasterMix2 to 2 ul each PCR product sample in a fresh 0.2 ml PCR tube.
- Vortex and pulse spin down.

3.7 Multiplex SBE Conditions

1. Program the thermocycler for the following conditions.

   96 °C for 2 min.

   25 cycles of 96 °C for 10 s and 50 °C for 5 s.

   60 °C for 30 s.

   15 °C forever.

3.8 Purification of the SBE Product

1. Add 1 μL SAP to each tube of PCR product.
2. Incubate in PCR machine.

   37 °C for 45 min.

   75 °C for 15 min.

   15 °C forever.

3.9 Capillary Electrophoresis

1. Mix 1 μL of cleaned SBE product with 8.7 μL HiDi Formamide and 0.3 μL Liz 120 size standard (per sample).
2. Vortex and pulse spin down.
3. Run mixture on AB Genetic Analyser (*see* **Note 9**) using POP-7 (*see* **Note 10**) on a 36 cm capillary length array and the following run parameters; injection voltage of 2.5 kV for 10 s and run time of 500 s at 60 °C.
4. Please ensure you are using the correct matrix for the SNaPshot chemistry (DS-02/Dye set E5).

3.10 Analysis Methods

3.10.1 HIrisPlex Genotype Scoring

1. Use suitable allele scoring software such as GeneMapper® or GeneMarker® when calling genotypes at each SNP. The details for each Bin can be found in Table 3 (*see* **Note 10**).
2. A 50 rfu peak height threshold should be used for all allelic calls. Ensure this by doing a manual check for each SNP (*see* **Note 11**).
3. Difficult samples (low concentration or possible drop out) should be run in duplicate for consistency of genotyping calls. For some troubleshooting examples, *see* **Note 12**.

3.10.2 HIrisPlex Prediction Model

1. Go to the website http://www.erasmusmc.nl/fmb/resources/Irisplex_HIrisPlex/
2. Click on the prediction webtool link and scroll to the HIrisPlex portion of the website. Input the allele results here according to the presence or absence of the named allele. You may input a full

Table 3
SNP Genotype peak colors and approximate bin size for capillary electrophoresis run on Applied Biosystems 3130xl, with analyses on the GeneMapper™ 4.0 software. Also included are the HIrisPlex genotypes for the 9947A positive control

SNP No.	SNP	SNPs	Peak color and approx. size (bp)	Peak color and approx. size	9947A positive control
1	N29insA	C/A	Black, 28–30	Green, 29–31	**CC**
2	rs11547464	G/A	Blue, 34–36	Green, 35–37	**GG**
3	rs885479	C/T	Black, 37–39	Red, 39–41	**CC**
4	rs1805008	C/T	Black, 40–42	Red, 41–43	**CC**
5	rs1805005	G/T	Blue, 43–45	Red, 45–47	**GG**
6	rs1805006	C/A	Black, 46–48	Green, 47–49	**CC**
7	rs1805007	C/T	Black, 48–50	Red, 50–52	**CC**
8	rs1805009	G/C	Blue, 52–54	Black, 52–54	**GG**
9	Y152OCH	C/A	Black, 54–56	Green, 54–56	**CC**
10	rs2228479	G/A	Blue, 56–58	Green, 57–59	**GG**
11	rs1110400	C/T	Black, 59–61	Red, 60–62	**TT**
12	rs28777	C/A	Black, 62–65	Red, 60–62	**AA**
13	rs16891982	G/C	Blue, 65–67	Black, 65–67	**GG**
14	rs12821256	G/A	Blue, 68–70	Green, 69–71	**AA**
15	rs4959270	C/A	Black, 71–73	Green, 72–74	**CA**
16	rs12203592	C/T	Black, 74–76	Red, 74–76	**CT**
17	rs1042602	G/T	Blue, 76–87	Red, 78–80	**GG**
18	rs1800407	G/A	Blue, 80–82	Green, 81–83	**GG**
19	rs2402130	G/A	Blue, 83–85	Green, 84–86	**AA**
20	rs12913832	C/T	Black, 87–89	Red, 88–90	**CC**
21	rs2378249	C/T	Black, 89–91	Red, 90–92	**CT**
22	rs12896399	G/T	Blue, 93–95	Red, 94–96	**TT**
23	rs1393350	C/T	Black, 95–97	Red, 96–98	**TT**
24	rs683	G/T	Blue, 98–100	Red, 99–101	**TT**

24 SNP profile or a partial profile (*see* **Note 14**) to generate eye and hair color predictions for a single individual. You can also predict several individuals at once using the file upload option.

3. Probability values are output for all three categories of eye color; blue, intermediate, brown, and hair color; blond, brown, red, black including hair shade categories of light and dark.

3.10.3 Understanding the HIrisPlex Prediction Probability Values

1. For eye color, the category with the highest probability value is defined as the most probable eye color prediction (*see* **Note 15**).
2. For hair color, *see* Fig. 1 for a guide on understanding the probability values, using the probability stepwise approach and producing the most probable hair color prediction (*see* **Note 15**).
3. Figure 2 gives both eye and hair color final prediction examples.

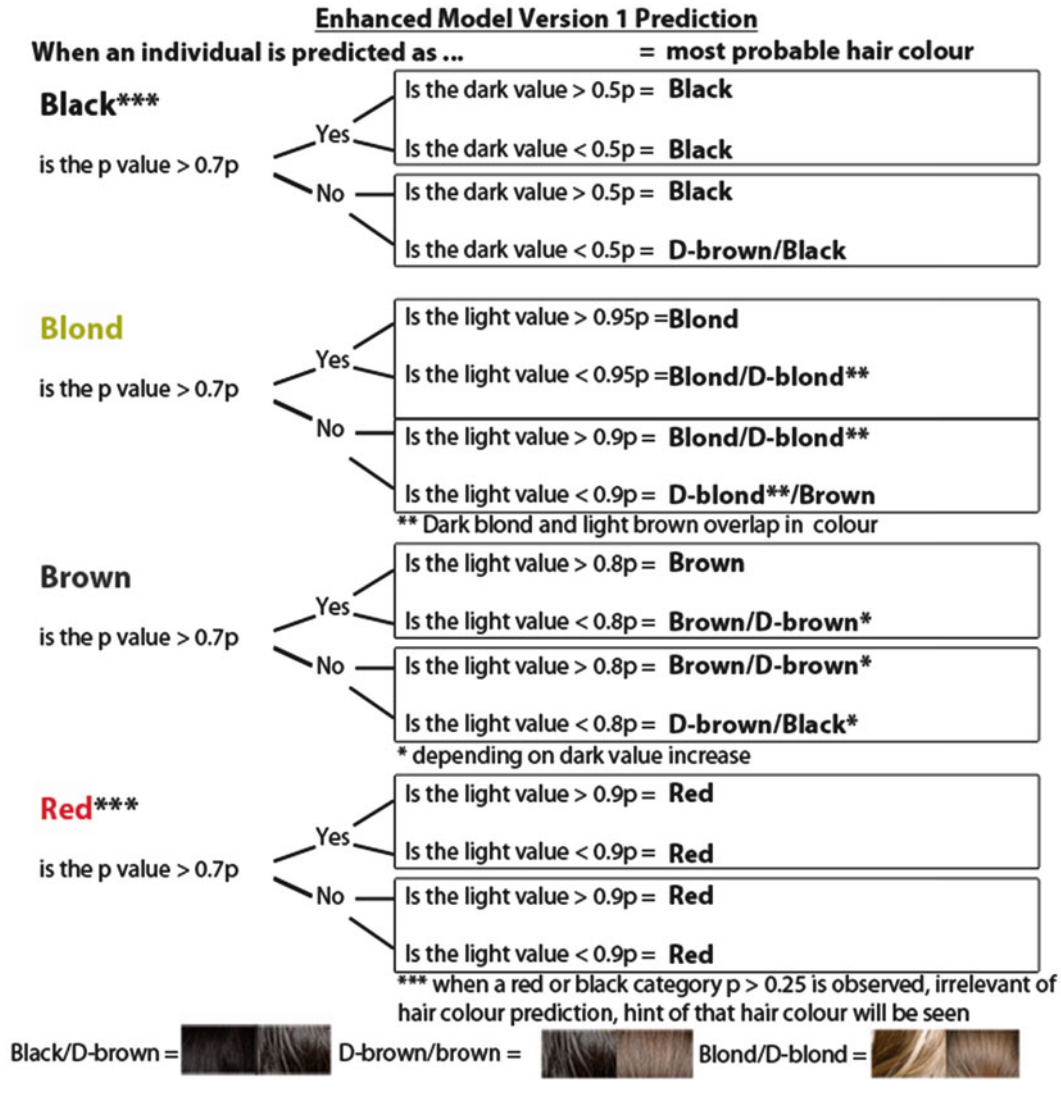
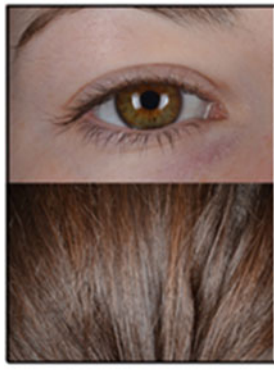


Fig. 1 HIrisPlex prediction guide (Walsh 2014) on how to interpret individual hair color and hair shade probabilities as derived from the HIrisPlex prediction tool available from Walsh et al. 2014 or online at http://www.erasmusmc.nl/fmb/resources/Irisplex_HIrisPlex/. D-Brown stands for dark brown and D-Blond stands for dark blond

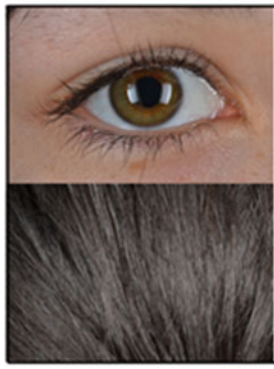


	Eye Colour (p)
Blue	0.35
Intermed.	0.228
Brown	0.422
	Hair Colour (p)
Blond	0.14
Brown	0.67
Black	0.186
Red	0.004
	Hair Shade (p)
Light	0.511
Dark	0.489

Final prediction:

Brown eye colour
Brown hair colour

Eye colour probabilities are very similar & lower than 0.5p. but highest is brown. Suggest a brown prediction but possible intermediate due to similar p values. Hair colour shows brown with second highest p value of black but at less than 0.25 p. Similar shades indicate light and dark even out, therefore should be brown

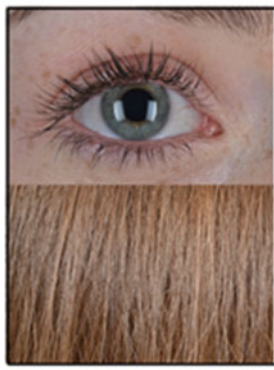


	Eye Colour (p)
Blue	0.036
Intermed.	0.111
Brown	0.853
	Hair Colour (p)
Blond	0.031
Brown	0.394
Black	0.574
Red	0.001
	Hair Shade (p)
Light	0.08
Dark	0.92

Final prediction:
***admixed individual**

Brown eye colour
Black hair colour

Eye colour highest probability of brown is very high at > 0.85 p, no other prediction Hair colour shows black but at a lower p value at 0.57 p. Next category is brown with with no significant blond/red..High dark shade also indicates black over dark brown

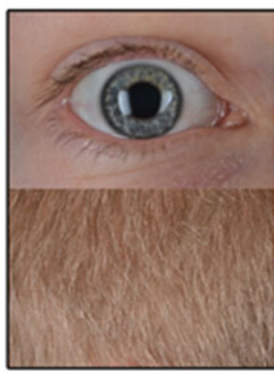


	Eye Colour (p)
Blue	0.946
Intermed.	0.043
Brown	0.011
	Hair Colour (p)
Blond	0.239
Brown	0.259
Black	0.016
Red	0.486
	Hair Shade (p)
Light	0.959
Dark	0.041

Final prediction:

Blue eye colour
Red (strawberry blond) hair colour

Eye colour highest probability of blue is near complete prediction of 1. No other colour in this prediction. Hair colour shows red but at a low p value with even blond and brown contributions and no black, therefore it can be red blond/red auburn but not dark red.



	Eye Colour (p)
Blue	0.952
Intermed.	0.04
Brown	0.008
	Hair Colour (p)
Blond	0.634
Brown	0.164
Black	0.005
Red	0.197
	Hair Shade (p)
Light	0.994
Dark	0.006

Final prediction:

Blue eye colour
Blond (dark-blond) hair colour

Eye colour highest probability of blue is near complete prediction of 1. No other colour in this prediction. Hair colour shows blond but at a lower p value than 7, this could indicate a blond/ dark blond, and there is red & brown at lower levels, more likely a darker blond.

Fig. 2 Final prediction examples using the HIrisPlex system (Walsh 2014) of four different individuals. The figure provides HIrisPlex probability values generated from each individuals DNA and includes a written interpretation of their genotype results by an experienced HIrisPlex user utilizing the HIrisPlex prediction guide (Walsh 2014). The figure also shows the individuals actual eye and hair color phenotype images for references

4 Notes

1. PCR primers were purchased from Metabion (Martinsried, Germany), desalted, and lyophilized. SBE primers were HPLC purified and lyophilized; 100 μM stock solutions were made with 1× TE buffer and stored at −20 °C. Further dilutions of working solutions should be made with ddH_2O and stored in the fridge at 4 °C when in use.
2. In previous developmental validation sensitivity assessments, the HIrisPlex assay produced full consistent profiles at a minimum DNA input of 64 pg. It is possible to produce full/partial profiles lower than this level; however its accuracy cannot be guaranteed (possible allelic drop out/in). For such low quantities, duplication of the profile should be performed for reliable consistent results in genotyping. In certain circumstances, splitting the assay into two, primers 1–12 in one PCR reaction and primers 13–24 in another first PCR reaction, then joining the resulting products together for the subsequent SBE reaction may be preferred for maximal use of low concentration template DNA.
3. The HIrisPlex system should not be used to differentiate between individuals within mixtures (as with all SNP typing); therefore samples where an eye and hair color prediction is required should be performed on single source samples. Prior STR profiling can ascertain this.
4. The design of the HIrisPlex system caters for highly degraded DNA down to fragments of approximately 150 bp and full profiles are expected in template DNA above this length. However partial profiles are possible with extremely degraded DNA, i.e., ancient DNA [6].
5. When making up working solutions, please follow Table 1 for concentrations for both PCR and SBE reactions and easy dispensing. If there is a volume noted under a certain solution concentration, make the primer in this working solution and take the noted volume for input into the master mix, i.e., for SNP1 the PCR primer working solution concentrations are 50 μM, taking 0.1 μL of the forward with 1 μL of the reverse for the reaction at a final primer concentration of 0.5 μM each in a 10 μL reaction volume. For the SBE reaction a 50 μM primer working solution concentration is made and 0.13 μL is put into a reaction for a final primer concentration of 1.3 μM. In the case of SNP2, the SBE primer working solution concentration is only 10 μM with 0.05 μL volume for a final reaction concentration of 0.1 μM in a 5 μL reaction volume.
6. The PCR negative control should not contain DNA, and ddH_2O should replace that volume. The PCR positive control

should contain a DNA sample that has previously worked before and which the profile is known, i.e., 9947A control DNA (Applied Biosystems). This profile can be found in Table 3.

7. The PCR reaction volume is 10 μL; however only 5 μL is required when cleaning up the PCR product with 2 μL ExoSAP-IT®. Therefore 5 μL should be transferred to fresh 0.2 ml tube. The remaining 5 μL can be stored at −20 °C for repeated profiling at a later date if required. In our experience, ExoSAP-IT® (USB Corporation) gives clearer and more sensitive results than using an in-house Exonuclease and SAP cleanup mixture.
8. It is required to run SNaPshot® Multiplex Chemistry with the manufacturers provided positive and negative controls to ensure efficient SBE amplification. It is also useful as a guide when troubleshooting to indicate at which point the reaction ran into difficulties, i.e., the first PCR reaction or the subsequent SBE reaction.
9. Although an ABI 3130*xl* capillary electrophoresis machine (Life Technologies) was used during the HIrisPlex System Developmental Validation, it is possible to use other ABI machines such as the newer ABI 3500. To note: this machine has been shown to provide increased peak heights and overall sensitivity; however fragment lengths differ slightly on the electropherogram.
10. POP7 polymer was also used for HIrisPlex validation purposes and it has been noted that differences in fragment length (especially in regions up to 40 bp in size) occur using POP4 but do not significantly affect the HIrisPlex systems overall performance.
11. Bins for each locus may not represent precisely the bins provided in Table 3, it depends on the machine and polymer; however these ranges should be used as a reliable guide and positive control 9947A can be found in the table to use as a size reference.
12. Standard STR profiling calls for a 50 relative fluorescence unit (rfu) threshold to be used to assess allele calls in a profile. The HIrisPlex system also uses this threshold; therefore allele peaks above are called and peaks below this level are not called.
13. Some troubleshooting examples:
    (a) There is no profile in your sample including your positive control; however the SNaPshot® positive control did produce a profile; there is an issue with the first PCR and its reagents, try repeating the first PCR.
    (b) There is no profile in your sample but positive control worked and the SBE SNaPshot® positive control did pro-

duce a profile; there is an issue with the first PCR and your DNA sample, check and quant your DNA sample.

(c) There is no profile in any of your samples including the SBE SNaPshot® positive control; there is an issue with the SBE reaction and the SNaPshot® mix, repeat using fresh SNaPshot® reaction mix.

(d) There are peaks in the first PCR negative control sample but not in the SBE positive control sample; there is possible contamination of your first PCR reagents.

(e) There are peaks in the PCR negative control and the SBE negative control; there is possible contamination of your SBE reagents.

(f) There are large peaks overlapping the HIrisPlex peaks at the beginning of the electropherogram up to 35 bp in size and they decrease peak heights of all the HIrisPlex alleles. This is an issue with the cleanup procedure of the first PCR using ExoSAP. Repeat cleanup with fresh ExoSAP and the extra available 5 μL PCR product that was stored. There may be small artifact peaks observed after 150 bp in size but they do not interfere with the HIrisPlex profile.

(g) Profile shows upward slope low to high. This can occur if reagents and primers from the first PCR reaction or the SBE reaction (such as SNaPshot®) are not optimal, i.e., freeze/thaw too many times. The primer sets (SBE primers 13–24) only contain a single SNP for amplification using the SBE primers; however the *MC1R* variants (primers 1–12) only have 4 PCR products and will therefore be the first affected if a PCR is not optimal. Repeat with fresh reagents.

(h) Very low profile peaks. Ensure all reagents and primers are fresh, also check DNA concentration. If it is a very low concentration (<64 pg) or ancient DNA sample, it may be wise to split the first PCR reaction into two. One reaction containing the *MC1R* PCR primers and the second containing the 13–24 primers. The PCR products can then be joined for the SBE reaction; this provides additional PCR product for the downstream reactions and may provide higher peak heights and less incidences of drop out. If the sample is extremely bad and HIrisPlex does not produce enough peaks, then IrisPlex (6-SNP assay) may be performed on the samples for eye color prediction alone.

14. If a full HIrisPlex profile was produced, follow the website in terms of what values you input for each locus. For example, for DNA variant 20 which is *HERC2* rs12913832, if the genotype at that locus was GG; input 0, if GT; input 1, if TT; input 2. If a partial profile is produced, then please input NA for that locus.

15. The HIrisPlex system is very accurate at predicting blue and brown eye color over the 0.7p probability threshold level and accurate over the 0.5p probability threshold level (see accuracy threshold levels in Walsh et al. 2012) [31]. However currently, predictions that have very similar probability values (<0.5p) are more difficult to predict and have an increased chance of error. This includes the intermediate category, i.e., green for which the HIrisPlex has the least level of accuracy in prediction, as the genes/SNPs associated with this trait are not believed to have been found yet. Further research is required to improve intermediate eye color category prediction.

For categorical hair color prediction the HIrisPlex system is currently the only tool available, with precision accuracy values (using the previous HIrisPlex model [29]) of 69.5 % for blond, 78.5 % for brown, 80 % for red, and 87.5 % for black hair on a >300 individual test set. A word of caution must be noted with blond hair color predictions; although the HirisPlex system is accurate at predicting blond hair, there are instances (in approx. 8 % of cases in a 119 test) where an individual has naturally changed hair color from childhood to adulthood (i.e., blond to dark brown/black) and these "age-dependent hair color changes" have not yet been implemented into the hair color prediction model. This means that a light blond hair color prediction could reflect a blond haired individual who has been blond all their life or a brown/black haired adult that had blond hair as a child and it changed during adolescence (ages vary). This error rate has already been implemented into the 69.5 % precision accuracy.

Acknowledgements

We thank Lakshmi Chaitanya for valuable comments. Initial work on the HIrisPlex system was funded in part by the Netherlands Forensic Institute (NFI), Erasmus MC University Medical Center Rotterdam, and by a grant from the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands (FGCN).

References

1. Tully G (2007) Genotype versus phenotype: human pigmentation. Forensic Sci Int Genet 1:105–110
2. Kayser M, Schneider PM (2009) DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. Forensic Sci Int Genet 3:154–161
3. Kayser M, de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet 12:179–192

4. Olalde I, Allentoft ME, Sanchez-Quinto F, Santpere G, Chiang CWK, DeGiorgio M et al (2014) Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature 507:225–228
5. King TE, Fortes GG, Balaresque P, Thomas MG, Balding D, Delser PM et al (2014) Identification of the remains of King Richard III. Nat Commun 5:5631. doi:10.1038/ncomms6631
6. Draus-Barini J, Walsh S, Pospiech E, Kupiec T, Glab H, Branicki W et al (2013) Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. Investig Genet 4:3. doi:10.1186/2041-2223-4-3
7. Brand A (2009) Integrative genomics, personal-genome tests and personalized healthcare: the future is being built today. Eur J Hum Genet 17:977–978
8. Janssens ACJW, van Duijn CM (2008) Genome-based prediction of common diseases: advances and prospects. Hum Mol Genet 17:R166–R173
9. Sulem P, Gudbjartsson D, Stacey S, Helgason A, Rafnar T, Magnusson K et al (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet 39:1443–1452
10. Duffy DL, Montgomery GW, Chen W, Zhao ZZ, Le L, James MR et al (2007) A three single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. Am J Hum Genet 80: 241–252
11. Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A et al (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet 4, e1000074
12. Branicki W, Brudnik U, Kupiec T, Wolańska-Nowak P, Szczerbińska A, Wojas-Pelc A (2008) Association of polymorphic sites in the OCA2 gene with eye colour using the tree scanning method. Ann Hum Genet 72:184–192
13. Sturm RA, Larsson M (2009) Genetics of human iris colour and patterns. Pigment Cell Melanoma Res 22:544–562
14. Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer K et al (2008) Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the *HERC2* gene inhibiting *OCA2* expression. Hum Genet 123:177–187
15. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M et al (2008) Two newly identified genetic determinants of pigmentation in Europeans. Nat Genet 40:835–837
16. Kayser M, Liu F, Janssens ACJW, Rivadeneira F, Lao O, van Duijn K et al (2008) Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. Am J Hum Genet 82:801
17. Mengel-From J, Wong T, Morling N, Rees J, Jackson I (2009) Genetic determinants of hair and eye colours in the Scottish and Danish populations. BMC Genet 10:88
18. Valenzuela RK, Henderson MS, Walsh MH, Garrison NA, Kelch JT, Cohen-Barak O et al (2010) Predicting phenotype from genotype: normal pigmentation. J Forensic Sci 55:315–322
19. Valverde P, Healy E, Jackson I, Rees JL, Thody AJ (1995) Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. Nat Genet 11:328–330
20. Kanetsky PA, Swoyer J, Panossian S, Holmes R, Guerry D, Rebbeck TR (2002) A polymorphism in the agouti signaling protein gene is associated with human pigmentation. Am J Hum Genet 70:770–775
21. Graf J, Hodgson R, van Daal A (2005) Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. Hum Mutat 25:278–284
22. Branicki W, Brudnik U, Draus-Barini J, Kupiec T, Wojas-Pelc A (2008) Association of the SLC45A2 gene with physiological human hair colour variation. J Hum Genet 53:966–971
23. Shekar SN, Duffy DL, Frudakis T, Sturm RA, Zhao ZZ, Montgomery GW et al (2008) Linkage and association analysis of spectrophotometrically quantified hair color in Australian adolescents: the effect of OCA2 and HERC2. J Invest Dermatol 128:2807–2814
24. Liu F, van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, Janssens ACJW et al (2009) Eye color and the prediction of complex phenotypes from genotypes. Curr Biol 19:R192–R193
25. Walsh S, Liu F, Ballantyne K, van Oven M, Lao O, Kayser M (2011) IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet 5:170–180
26. Ruiz Y, Phillips C, Gomez-Tato A, Alvarez-Dios J, Casares de Cal M, Cruz R et al (2013) Further development of forensic eye color predictive tests. Forensic Sci Int Genet 7:28–40

27. Hart KL, Kimura SL, Mushailov V, Budimlija ZM, Prinz M, Wurmbach E (2013) Improved eye- and skin-color prediction based on 8 SNPs. Croat Med J 54:248–256
28. Branicki W, Liu F, van Duijn K, Draus-Barini J, Pośpiech E, Walsh S et al (2011) Model-based prediction of human hair color using DNA variants. Hum Genet 129: 443–454
29. Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A et al (2013) The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int Genet 7:98–115
30. Walsh S, Chaitanya L, Clarisse L, Wirken L, Draus-Barini J, Kovatsi L et al (2014) Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci Int Genet 9:150–161
31. Walsh S, Wollstein A, Liu F, Chakravarthy U, Rahu M, Seland J et al (2012) DNA-based eye colour prediction across Europe with the IrisPlex system. Forensic Sci Int Genet 6:330–340
32. Pospiech E, Draus-Barini J, Kupiec T, Wojas-Pelc A, Branicki W (2011) Gene-gene interactions contribute to eye colour variation in humans. J Hum Genet 56:447–455
33. Pośpiech E, Wojas-Pelc A, Walsh S, Liu F, Maeda H, Ishikawa T et al (2014) The common occurrence of epistasis in the determination of human pigmentation and its impact on DNA-based pigmentation phenotype prediction. Forensic Sci Int Genet
34. Liu F, Wollstein A, Hysi PG, Ankra-Badu GA, Spector TD, Park D et al (2010) Digital quantification of human eye color highlights genetic association of three new loci. PLoS Genet 6, e1000934
35. Andersen JD, Johansen P, Harder S, Christoffersen SR, Delgado MC, Henriksen ST et al (2013) Genetic analyses of the human eye colours using a novel objective method for eye colour classification. Forensic Sci Int Genet 7:508–515

Chapter 18

Inference of Ancestry in Forensic Analysis I: Autosomal Ancestry-Informative Marker Sets

Chris Phillips, Carla Santos, Manuel Fondevila, Ángel Carracedo, and Maria Victoria Lareu

Abstract

An expanding choice of ancestry-informative marker single nucleotide polymorphisms (AIM-SNPs) is becoming available for the forensic user in the form of sensitive SNaPshot-based tests or in alternative single-base extension genotyping systems (e.g., Sequenom iPLEX) that can be adapted for analysis with SNaPshot. In addition, alternative ancestry-informative variation: Indels and STRs can be analyzed using direct PCR-to-CE techniques that offer the possibility to detect mixed profiles. We review the current forensically viable AIM panels, their optimized PCR multiplexes, and the population differentiation power they offer. We also describe how improved population divergence balance can be achieved with the enlarged multiplex scales of next-generation sequencing approaches to enable analysis of admixed individuals without biased estimation of co-ancestry proportions.

Key words Genetic ancestry, SNP, AIM-SNP, SNaPshot, Indels, Population genetics, Next-generation sequencing (NGS)

1 Introduction

The inference of ancestry in forensic analysis, by typing DNA variants with highly differentiated allele frequencies amongst populations, has the scope to provide a reliable genetic-based substitute for eyewitness when this is not available to investigators [1–3]. In fact, the analysis of genetic variation to infer ancestry has a wide range of applications in forensic testing beyond attempts to gain information about unknown donors of DNA found at the crime scene, including (1) achieving a more complete identification of skeletal remains from missing persons or mass grave sites in regions of conflict; (2) corroborating eyewitness accounts of the perceived

Electronic supplementary material: The online version of this chapter (doi:10.1007/978-1-4939-3597-0_18) contains supplementary material, which is available to authorized users.

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_18,

ethnicity of a person when these are unreliable or conflicting (*see* **Note 1**); (3) confirming the self-declared ancestry of donors collected for genetic databases such as those compiled for STRs, Y-markers, and mitochondrial variation (the latter two showing marked geographic differentiation so that checking the ancestry of outliers is critical to database accuracy); (4) refining familial search strategies that are highly dependent on STR allele frequency assumptions made prior to the search [4]; (5) helping assess atypical combinations of physical characteristics (using, e.g., Irisplex [5]) due to admixed ancestry in the individual; and (6) enhancing studies where forensic-level sensitivity is necessary such as testing medical archive material or archaeological remains.

In the previous publication on forensic ancestry testing in Methods in Molecular Biology 830 [6], we outlined ways to type forensic material and analyze the resulting allelic data with a SNaPshot single-base extension (SBE) assay of 34 ancestry-informative marker (AIM) SNPs [1]. Subsequently the 34-plex AIM-SNP assay was enhanced by swapping out a poorly performing SNP and completing more comprehensive control DNA typing to add reference data for first-time SNaPshot users [7] while online SNP frequency databases have continued to grow in both depth and scope. A complimentary SNaPshot forensic ancestry test has also been developed to use alongside 34-plex termed *Eurasiaplex*: designed to differentiate South Asians from Europeans [8]. Potentially more useful for forensic geneticists interested in analyzing ancestry with tests as straightforward as routine DNA profiling, SNaPshot tests have now also been extended with ancestry-informative insertion-deletion marker sets (AIM-Indels) [9, 10] and ancestry-informative STR (AIM-STR) sets [11]. Since both marker sets are typed using dye-labeled PCR primers and tests place PCR amplification products directly into capillary electrophoresis (direct PCR-to-CE), the peak balance obtained is much better than SNaPshot such that electrophoretic signals more directly reflect the allelic ratios of input DNA. This characteristic markedly improves profile quality and introduces the possibility of mixture detection regimes similar to those routinely applied to kit-based STR analysis. Therefore, forensic laboratories have a range of multiplexes to choose from that can help build a picture of the ancestry of DNA from unknown sources. In parallel to an expanded range of forensic ancestry tests, the online Bayesian classifier *Snipper* has undergone a series of enhancements to its functionality (http://mathgene.usc.es/snipper/index.php). We have divided the description of current forensic ancestry analysis into two complimentary chapters covering here: AIM marker choice and forensically applicable multiplexes, and in the next chapter: the options available for analysis of the genetic data generated from these tests.

Lastly, important developments have been gathering pace in the field of next-generation sequencing (NGS) analysis. The

possibility to sequence short DNA fragments amplified in PCR multiplexes of several hundred markers with bench-top scale systems has obvious implications for expanding the scope of forensic DNA analysis. NGS systems offer the real possibility to combine in a single test: identification markers, ancestry-informative markers, and SNPs able to predict common physical characteristics such as eye and hair color. Even if just a proportion of the multiplexing capacity is used for ancestry markers, then 100–200 sites can be combined to significantly improve the data depth for inference of ancestry with raised precision. The question then becomes: what AIMs should be used to make best use of this capacity and what additional characteristics might be sought for the set as a whole? We will explore the concept of population differentiation balance as a way to analyze population admixture with reduced bias when estimating co-ancestry proportions. This factor was used to reconstruct a new forensic AIM-SNP set of various multiplex scales specifically for NGS platforms.

2 Materials

Previously provided URLs to access genetic data online are unchanged [6] but database scope has expanded. Two important additions relevant to forensic SNP analysis are the FROGkb and Complete Genomics databases.

1. The *Snipper* options can be accessed from a menu of eight well-defined analytical approaches at: http://mathgene.usc.es/snipper/. Fixed training sets have now been extended to 34 SNPs, 46 Indels, and their combination into 80-marker profiles. Our recommendation for appropriate population scope when using AIM set combinations is to choose three-group classifications of Africans, Europeans, East Asians in *Snipper* with 34-plex, four (adding Americans) with Indels, and five (adding Oceanians) when combining both. Cross-validation with *Snipper* is a commonly used population data and SNP power assessment routine and is selected at: http://mathgene.usc.es/snipper/analysispopfile2_new.html
2. SPSmart similarly offers all database choices from one page (at: http://spsmart.cesga.es). Accessible data includes the full 1000 Genomes variant catalog (ENGINES), 70 forensic STRs (pop.STR), SNP*for*ID forensic SNP set data and HapMap, Perlegen, and Stanford/Michigan HGDP-CEPH SNP catalogs (*see* **Note 2**).
3. FROGkb (at: http://frog.med.yale.edu/FrogKB/) is the Forensic Reference Resource On Genetics run by Kiddlab. FROGkb lists several forensic SNP sets for ancestry analysis, identification, and prediction of physical characteristics. The

site also includes a likelihood calculator for each SNP set for a range of populations linked in from the ALFRED variant database (http://alfred.med.yale.edu).

4. Complete Genomics (CG) has public listing of variant data for 427 complete genomes (at: http://www.completegenomics.com/public-data/). All current CG genomes are from study individuals common to 1000 Genomes but, unlike 1000 Genomes, also include variant site data for mt-DNA, X and Y-chromosomes.

3 Methods

3.1 The Revised SNPforID 34-Plex Forensic Ancestry Test

1. A revision of the 34-plex SNaPshot test swapped the underperforming AIM-SNP rs727811 with the highly informative rs3827760 that has a near-fixed East Asian specific allele (*see* **Note 2**). Additionally, SBE primers of rs2304925, rs5997008, rs2814778, rs239031, and rs16891982 were modified to adjust peak positions when detecting genotypes with POP-4™ polymer (Life Technologies), while certain mobility modifying poly-CT tails were changed to nonhomologous sequences. This adaptation pre-empted increasing use of AB POP-4™ and likely discontinuation of AB POP-6™ polymer for which the original 34-plex assay was developed. We moved rs2304925/rs5997008 away from the fastest mobility size range where dye artifacts can interfere with allelic peak recognition and rs2814778/rs239031/rs16891982 were size-adjusted to better separate them from neighboring SNP peaks or nonspecific signals seen with POP-4™. Details of the 34-plex assay SNPs are summarized in Table 1.

3.1.1 PCR Amplification of 34-Plex

1. Make a master mix containing all primers diluted to the concentration shown in Supplementary File S1A.
2. Set up PCR reactions with 1× AB AmpliTaq Gold PCR buffer; 25 mM $MgCl_2$ to a final concentration of 5.9 mM; 10 mM dNTP mix to a final concentration of 0.58 mM; 3.2 mg/μl BSA to a final concentration of 0.29 mg/μl, 0.5 U of AB AmpliTaq Gold polymerase; 1 μl of premixed PCR primers; 0.75 ng of DNA in 6.9 μl final reaction volume.
3. PCR cycling: 95 °C for 15 min, then 30–32 cycles at 95 °C for 30 s, 60 °C for 50 s, and 65 °C for 40 s with a final extension at 65 °C for 15 min.
4. Make a master mix of SBE primers according to the details in Table S1.
5. To clean up PCR products add 1.3 μl of Illustra ExoProStar 1-step (GE Healthcare) to 2.5 μl of PCR product. Note this is a name change of ExoSAPit for licensing reasons and both products are identical.

Table 1
Summary of component marker characteristics of the revised 34-plex SNP-based ancestry test [7]

SNaPshot position	rs-number	USC code	SNaP Bases[a]	1000-G bases	Expt size	Obs bp: G	Obs bp: A	Obs bp: C	Obs bp: T	Het ratio
1	rs1321333	P03	TC	AG	22			26.89	28.63	1.31
2	rs917118	A07	GA	CT	24	28.17	30.16			2.34
3	rs1024116	A29	GA	CT	28	26.75	28.95			1.18
4	rs7897550	P05	CT	GA	28			31.44	32.87	1.19
5	rs722098	A21	AG	AG	33	33.88	36.28			3.53
6	rs10843344	P06a	CT	CT	35			37.05	38.31	1.36
7	rs239031	P07	CT	GA	36			38.9	40.45	1.17
8	rs12913832	P08	AG	AG	38	40.15	40.9			4.89
9	rs2040411	A40	GA	GA	41	43.18	44.25			1.56
10	rs1978806	P09a	TC	AG	41			44.27	45.52	1.21
11	rs773658	P10	GC	CG	45	46.25		46.73		3.10
12	rs10141763	P11	AT	TA	45		48.74		49.32	1.12
13	rs182549	P12	CT	CT	49			49.11	50.55	1.10
14	rs1573020	P13	GA	GA	49	50.27	50.98			1.10
15	rs896788	P14	CT	CT	53			52.52	53.26	1.21
16	rs2065160	P15	AG	AG	53	54.11	54.82			2.38
17	rs2572307	P16a	AG	AG	57	55.41	56.02			1.41
18	rs2303798	P17	CT	GA	57			57.14	57.73	2.19
19	rs2065982	P18	AG	TC	61	59.17	59.9			1.71
20	rs3785181	P19	CT	CT	61			60.27	61.49	1.54
21	rs881929	P20	GT	GT	64	61.83			63.86	2.07
22	rs1498444	P21	AC	TG	65		64.17	64.01		1.19
23	rs1426654	P22a	TC	AG	68			67.35	68.19	1.26
24	rs2026721	P23	AG	TC	69	68.45	69.37			2.04
25	rs4540055	P24	TCA	AC	73		72.15	71.68	72.48	2.16[b]
26	rs16891982	P25a	GC	CG	75	76.03		76.44		1.76
27	rs1335873	A52	TA	AT	77		78.16		78.31	0.99
28	rs1886510	A13	GA	GA	80	78.71	79.69			1.47
29	rs730570	P26	CT	GA	80			80.11	80.74	1.68
30	rs5030240	P27	GCA	CA	85	83.75	84.91	84.51		2.02[c]

(continued)

Table 1
(continued)

SNaPshot position	rs-number	USC code	SNaP Bases[a]	1000-G bases	Expt size	Obs bp: G	Obs bp: A	Obs bp: C	Obs bp: T	Het ratio
31	rs2304925	P01	TG	AC	87	86.61			87.93	2.32
32	rs5997008	P02	CA	CA	87	90	88.2	87.86		1.78
33	rs3827760	P28	AG	AG	90	89.64	90.09			5.08
34	rs2814778	P04	TC	TC	90			91.04	91.93	0.91

USC code refers to the internal name used for each marker; 1000-G = 1000 Genomes; expt/obs = expected/observed; bp base pair sizes; Het = heterozygote peak pair
[a]Reference allele first
[b]Listed value for P24-AC with P24-CT = 1.21 and P24-AT = 1.76
[c]Listed value for P27-AC with P27-AG = 1.64 and P27-CG = 3.57

6. For Single-Base extension, add 1 μl of cleaned PCR product, 1 μl of SNaPshot ready reaction mix, 0.5 μl of premixed SBE primers (ratios listed in Supplementary File S1A, previously 0.75 μl); 0.5 μl water.
7. Carry out the single-base extension using the following parameters: 28–30 cycles at 96 °C for 10 s, 55 °C for 5 s and 60 °C for 30 s.
8. In preparation for capillary electrophoresis (CE) combine 1 μl of a 1/25 dilution of SBE product with 8.9 μl of deionized Hi-Di formamide and 0.3 μl AB GS-120 LIZ size standard.
9. Run on an ABI PRISM 3130 Genetic Analyzer (AB), filter set G5, POP-4 polymer or equivalent (see NOTE—I'll add a note here to say that you can use other platforms, but the sizes and ht balances may change).

The changes to SNP positions and chemistry improve signal balance and aid more reliable analysis of SNaPshot peak patterns compared to the previous configuration of the 34-plex test. More importantly, the inclusion of rs3827760 has raised classification success rates for three of five population groups analyzed with cross-validation results indicating the following improvements: Europeans = 98.7 % raised to 99.4 %; East Asians = 92.5–94.7 %; Oceanians = 96.1–100 %. Africans and Americans remain at 100 % success with both SNP sets.

3.2 Eurasiaplex, a Complimentary Test to 34-Plex for Differentiating Europeans and South Asians

Eurasiaplex comprises 23 AIM-SNPs chosen to maximize the differentiation of European and South Asian populations [8]. The SNPs were also designed to be independent of 34-plex markers by being well separated in the genome; in this way European classifications with 34-plex could be further explored without bias due to linkage by supplementary *Eurasiaplex* analysis. A key aspect of the successful ancestry assignment of Europeans and South Asians was the application of a probability threshold of two orders of magnitude or "100 times more likely"—i.e., values below this minimum likelihood ratio of the most likely ancestry and the next most likely were not assigned (*see* **Note 3**).

1. Make PCR primer mix following the details in Table Supplementary File S1B.
2. Set up the PCR with 2 μl Qiagen multiplex PCR master mix; 2 μl primer mix; 1 ng of DNA in 5 μl final reaction volumes.
3. Carry out PCR: 95 °C for 15 min, then 35 cycles at 94 °C for 30 s, 60 °C for 60 s, and 72 °C for 50 s, with a final extension at 72 °C for 10 min.
4. Make PCR primer mix following the details in Table Supplementary File S1B.
5. Clean PCR products with 1.3 μl of Illustra ExoProStar 1-step (GE Healthcare) added to 2.5 μl of PCR product.
6. For SBE add 1 μl of cleaned PCR product, 1.25 μl of SNaPshot ready reaction mix, 0.75 μl of premixed SBE primers.
7. Carry out SBE cycling: 30 cycles at 96 °C for 10 s, 55 °C for 5 s, and 60 °C for 30 s.
8. For CE combine 1 μl of amplified product with 10 μl of deionized Hi-Di formamide and 0.25 μl AB GS-120 LIZ size standard, run on an ABI PRISM 3130 Genetic Analyzer (Life Technologies), filter set G5, POP-4 polymer.

3.3 The 46-Plex AIM-Indel Ancestry Test

Short Indels consisting of sequence insertions or deletions of between 1 and 25 bases make up ~5 % of human genome variants. Indels combine the favorable characteristics of SNPs with those of STRs by being easily typed from short amplified fragments, thus giving typing success from highly degraded DNA comparable to most SNPs, while their detection with dye-labeled PCR primers matches the proven genotyping system of forensic STRs [10]. The major advantage of direct PCR-to-CE is that a single reaction step restores the simple relationship between allele peak height ratios and input DNA—making mixture detection more reliable than when using SNaPshot [12]. The ancestry-informative Indels we identified were collected into a 46-plex assay designed to provide differentiation of Africans, Europeans, East Asians, and Native Americans [11].

1. Make primer mix with primers listed in Supplementary File S1C with all primers 1 mM.
2. Set up PCR with 5 μl Qiagen multiplex PCR master mix; 1 μl of the primer mix (i.e., final concentrations of 0.1 mM); 0.3–5 ng of DNA in 10 μl final reaction volume.
3. Carry out PCR: 95 °C for 15 min, then 28 cycles at 94 °C for 30 s, 60 °C for 90 s, and 72 °C for 60 s, with a final extension at 72 °C for 60 min.
4. Combine 0.8 μl of amplified product (although this may need dilution up to 1/10 with water) with 11.65 μl of deionized Hi-Di formamide and 0.35 μl AB GS-500 LIZ size standard for capillary electrophoresis.
5. Analyze samples by running on an ABI PRISM 3130 Genetic Analyzer, filter set G5, POP-4 polymer. Note that POP-7 was described in the original paper [11], but use of POP-4 is a viable alternative and the sizes listed in Supplementary File S1C are for POP-4 polymer.
6. Provisional analysis of Indel profiles with artificial mixtures indicates simple two-contributor mixed profiles can be readily detected as nonstandard peak patterns each time [12]. Figure 1 illustrates profiles of NED-labeled AIM-Indel products for NIST standard reference material DNA SRM2391-D, comprising a 1:3 combination of A:C DNA. Strong signal imbalances are evident in five peak pairs where contributor genotypes contrast and create skewed peak height ratios ranging from a minor peak height 65 % of the major peak, to as low as 15 %. In contrast, unmixed minor:major peak ratios average 94 % including the outlier rs2307803. Such patterns indicate simple

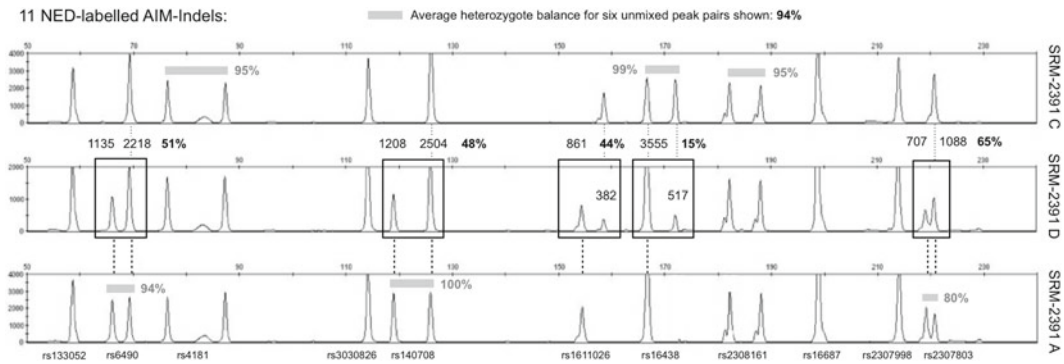


Fig. 1 NED-labeled AIM-Indel peaks obtained with standard control DNAs: SRM-2391-A, -C, and -D. Sample D is a 3:1 mixture of A:C so the *boxed peak pairs* in the *middle electropherogram* show varying degrees of imbalance as a result. A range of heterozygote peak pairs indicated in the *upper/lower electropherograms* suggest unmixed peak balance per locus averages ~94 % in contrast to the 15–65 % imbalance found in SRM 2391-D. Note that peak balance is primarily within-locus and signal strength is far from uniform between loci

mixed DNA is detectable using Indels particularly with 42 data points to permit comprehensive peak height comparisons. An important factor here is that mixtures of contributors from different ancestries will regularly show a larger number of genotype contrasts, so this type of mixture will be more readily detected. In fact, the SRM2391-D mixed DNA control comprises a combination of European and Oceanian donors. To further underline this aspect it is interesting to note that the apparent "Heterozygosity" (proportion of peak pairs) in the middle electropherogram of Fig. 1 is 64 %: much higher than an average proportion of 36 % seen in the other unmixed electropherograms.

An alternative 21 AIM-Indel multiplex has been developed by Zaumsegel et al. [13] that has no overlapping loci with the 46 Indels described above. This combination of 21 Indels has considerable free electrophoretic space as only 6FAM and HEX dye labels were used. Furthermore there is scope for re-balancing levels of differentiation obtained with these AIM sets from the combination of 46 and 21 Indels, as described in Subheading 3.6.

3.4 Inference of Ancestry from STR Data

Past initiatives to use forensic STRs for ancestry inference applied two different approaches: using Bayesian analysis to derive a likelihood ratio of most probable ancestry [14–16], or development of novel STR sets specifically to analyze ancestry [17]. The first approach using standard forensic identification STRs cannot adequately differentiate all populations, while the second has focused on di-nucleotide repeat STRs, inappropriate for forensic use due to very high stutter product peaks. Rosenberg et al. used much larger STR sets of 377 loci to successfully differentiate global populations [18] but multiplexes beyond 10–15 STRs are unsuitable for typing limited quantities of DNA. We used frequency data from the Rosenberg studies to identify the most differentiating tetra-nucleotide repeat STRs [18] to construct a single assay of ancestry-informative STRs (or AIM-STRs). In order to analyze STR profiles with frequency-based reference data the *Snipper* portal was adapted to handle reference allele frequencies in place of genotypes. Input of allele frequencies permits ancestry analysis with a flexible choice of marker type (e.g., combining STRs with SNPs) to allow reference-training sets to be easily constructed from the complex allelic data characteristic of STRs.

1. Make primer mix following details in Table Supplementary File S1.
2. Set up PCR: 1 μl 10× AB GeneAmp PCR buffer II; 0.8 μl 3.64 mM $MgCl_2$; 0.25 μl 0.23 mM dNTPs; 0.5 μl BSA (0.07 μg/μl); 0.2 μl AB AmpliTaq Gold*; 5 μl Qiagen Multiplex PCR Mix*; 1.82 μl PCR primer mix; 0.4–1.2 ng DNA in 12 μl final reaction volume. *We have found combining two different sources of polymerase can improve PCR performance.

3. Carry out PCR: 95 °C for 15 min, then 28–30 cycles at 96 °C for 30 s, 59 °C for 90 s, and 72 °C for 90 s, with a final extension at 72 °C for 60 min.
4. Combine 1 μl of amplified product with 9.5 μl of a 33:1 mixture of deionized Hi-Di formamide and AB GS-500 LIZ size standard, run on an ABI PRISM 3130 Genetic Analyzer, filter set G5, POP-7 polymer.

To analyze STR data for ancestry inference with *Snipper* the "Classification with a custom Excel file of frequencies" option is chosen (http://mathgene.usc.es/snipper/frequencies_new.html). Training sets for a range of commonly combined STR kits are provided for a total of 32 STRs (combining Promega Powerplex Fusion and Qiagen HDplex loci); 20 (Identifiler and NGM) or 12 AIM-STRs as described above [11]. Allele frequency data can also be collected from pop.STR by selecting population and STR combinations then collating tables of allele frequencies from the downloads page.

To reformat allele frequencies for input to *Snipper*, download the pop.STR csv files and open in Excel and the "Text to Columns" function (applying semicolon delimiters) converts all data to a table. STR profiles follow the input format: STR1Allele1,STR1Allele2/STR2Allele1,STR2Allele2/etc. and uses marker order: STR1 = worksheet1, STR2 = worksheet2, etc. Frequencies for STRs and SNPs can be combined but this follows the rather clumsy format of a different worksheet for each SNP with frequency data as "alleles × populations" tables.

Although 20 standard identification STRs are not as differentiating as the best AIM-SNPs, they still provide data that can be used towards an overall picture of an individual's ancestry. However the combination of the recently introduced Promega Powerplex Fusion and Qiagen Investigator HDplex STR kits provides 32 unique STRs from just two kits. The genotype data obtained from this number of STRs is certain to provide better differentiation of populations and is worth exploring further to gauge how well ancestry inferences match the population of origin of test samples. The frequency-based *Snipper* classifier now includes a training set for the above 32 STRs collating data from all seven population groups of the HGDP-CEPH panel.

3.5 Alternative SNP-Based Forensic Ancestry Tests

Alternative autosomal SNP ancestry tests using SNaPshot have been described by Lao et al. [19] and Gettings et al. [20]. As Lao's tests use relatively small-scale 12-plex amplifications, they are likely to provide similar levels of forensic sensitivity to the tests described above. Furthermore, there is only one SNP in common between these 24 and the 34-plex components: rs16891982. Therefore, including the 46-plex AIM-Indel test, up to 103 unique AIMs can be genotyped with four proven forensic tests to achieve considerable data depth for differentiating all five principal population groups

of Africans, Europeans, East Asians, Americans, and Oceanians. The details of the SNaPshot tests developed by Lao were outlined in the Supplementary Methods file of [18]. In a slightly different approach tests developed by Gettings combine AIMs and eye color predictive SNPs together [20]. While this is a worthwhile feature, the ancestry inference performance is compromised to some degree by including SNPs weakly predictive for pigmentation that have low levels of population differentiation (*see* **Note 1**). Furthermore, the SNPs are combined into three separate multiplexes, although data from 32 AIM-SNPs can be applied separately for ancestry analysis (see listing in FROGkb) and the paper includes *Snipper* training sets for this purpose (US populations of: Europeans, African Americans, Hispanics, and East Asians) as supplementary files [20]. Lastly, the FROGkb website lists a set of 55 candidate AIM SNPs, of which 41 have been developed as iPLEX assays (the Sequenom spectrometric SBE-based genotyping system) described by Nievergelt et al. [21]. The iPLEX assays would be easily adaptable to SNaPshot chemistry and these loci are likely to be of interest for laboratories aiming to build new ancestry tests or combine the best markers from each of four optimized forensic tests.

Individual PSD values in the three main population groups are listed in Table 2 for the four forensic AIM-SNP sets described in Subheading 3.1 and above. There is limited commonality between the four sets accounting for ~13 % of chosen markers. These common AIMs in multiple sets are listed in the top left sections of Table 1. In summary, common SNPs comprise rs16891982, the only marker present in all sets; rs3827760, rs2814778 plus pigmentation loci rs1426654 and rs12913832 in 34-plex, Gettings and FROGkb; rs1876482 in Lao, Gettings and FROGkb—a non-coding SNP commonly selected as a powerful E Asian-informative marker. Another 11 AIMs are present in two sets.

3.6 Redesigning Forensic Ancestry Sets for NGS: Selecting AIM-SNPs for Power and Population Divergence Balance

Small-scale forensic AIM sets need to collect the most powerful markers possible, as tests must generate the best ancestry-indicative data from limited amounts of DNA. For this reason, whole-genome scan approaches that can type up to a million markers but require large quantities of input DNA have never gained traction for forensic analysis despite the promise of much greater SNP data depth and therefore a higher level of geographic resolution compared to SNaPshot. NGS systems however can bridge this gap by genotyping up to 400 SNPs (and Indels or STRs) by direct sequence analysis with indications of equivalent or better sensitivity than SNaPshot.

To rebuild a forensic ancestry set for NGS analysis, we prioritized two properties of the SNP set: maximum differentiation power, as defined by the allele frequencies in the populations a test seeks to compare, and divergence balance achieved by obtaining near-identical differentiation in all population comparisons. If one or more differentiations are stronger than the others, then

Table 2
Individual African, European, and East Asian PSDs in 134 AIM-SNPs combined in four forensic ancestry tests (panels 1–4 are described in [7, 19–21] respectively)

Common AIMs	AFR	EUR	E ASN	2. Lao AIMs	AFR	EUR	E ASN	4. FROGkb AIMs	AFR	EUR	E ASN
A: rs16891982	0.230	0.588	0.272	rs1448484	0.466	0.121	0.105	rs1871534	0.502	0.131	0.108
B: rs3827760	0.150	0.204	0.477	rs1369290	0.458	0.091	0.116	rs310644	0.493	0.079	0.095
rs2814778	0.675	0.152	0.124	rs1478785	0.374	0.029	0.080	rs9522149	0.111	0.328	0.213
rs1426654	0.187	0.618	0.297	rs1371048	0.113	0.047	0.301	rs2196051	0.133	0.322	0.197
rs12913832	0.159	0.344	0.197	rs1461227	0.002	0.197	0.221	rs3916235	0.422	0.086	0.111
C: rs1876482	0.140	0.102	0.294	rs1405467	0.297	0.011	0.111	rs7326934	0.418	0.093	0.097
D: rs896788	0.002	0.060	0.085	rs1907702	0.293	0.067	0.011	rs4918664	0.163	0.098	0.347
rs730570	0.057	0.212	0.105	rs2052760	0.241	0.016	0.086	rs1229984	0.122	0.148	0.332
rs2065982	0.075	0.123	0.284	rs721352	0.149	0.000	0.172	rs3823159	0.395	0.136	0.031
E: rs952718	0.229	0.045	0.047	rs1858465	0.208	0.079	0.003	rs917115	0.045	0.273	0.215
rs722869	0.065	0.123	0.300	rs1048610	0.068	0.030	0.183	rs2238151	0.164	0.256	0.110
rs714857	0.133	0.224	0.045	rs1391681	0.149	0.000	0.098	rs10497191	0.397	0.064	0.057
rs1344870	0.011	0.036	0.063	rs1465648	0.185	0.033	0.018	rs3811801	0.093	0.144	0.275
F: rs4891825	0.394	0.049	0.132	rs2179967	0.143	0.056	0.006	rs1462906	0.338	0.081	0.087
rs3737576	0.020	0.000	0.007	rs926774	0.001	0.031	0.042	rs7251928	0.339	0.013	0.152
rs260690	0.021	0.319	0.270	rs1667751	0.051	0.008	0.002	rs2593595	0.389	0.036	0.066
rs1800414	0.091	0.140	0.266	rs1808089	0.009	0.010	0.001	rs4833103	0.114	0.234	0.141
				3. Gettings AIMs	*AFR*	*EUR*	*E ASN*	rs3814134	0.295	0.191	0.000

(continued)

1. 34-plex AIMs	*AFR*	*EUR*	*E ASN*	rs1834640	0.208	0.557	0.214	rs735480	0.325	0.158	0.000
rs881929	0.325	0.003	0.259	rs1375164	0.194	0.291	0.105	rs1834619	0.127	0.090	0.256
rs773658	0.407	0.093	0.082	rs2714758	0.412	0.089	0.083	rs7657799	0.310	0.143	0.010
rs239031	0.339	0.106	0.054	rs4891825	0.394	0.049	0.132	rs7554936	0.330	0.012	0.090
rs1335873	0.406	0.028	0.014	rs10007810	0.409	0.038	0.078	rs11652805	0.339	0.063	0.027
rs1573020	0.328	0.065	0.041	rs10496971	0.097	0.119	0.306	rs1572018	0.234	0.178	0.015
rs4540055	0.251	0.096	0.033	rs3784230	0.302	0.001	0.154	rs6754311	0.113	0.202	0.108
rs2026721	0.296	0.016	0.048	rs1667394	0.180	0.211	0.047	rs7997709	0.024	0.138	0.237
rs3785181	0.096	0.141	0.118	rs916977	0.178	0.212	0.047	rs798443	0.291	0.094	0.003
rs2040411	0.170	0.026	0.153	rs714857	0.133	0.224	0.045	rs7226659	0.070	0.087	0.193
rs2303798	0.195	0.146	0.001	rs6548616	0.306	0.013	0.081	rs6990312	0.206	0.014	0.100
rs2572307	0.230	0.033	0.058	rs885479	0.094	0.071	0.205	rs200354	0.035	0.161	0.086
rs1978806	0.144	0.068	0.101	rs7170852	0.106	0.185	0.054	rs192655	0.003	0.143	0.123
rs1321333	0.094	0.145	0.069	rs4778138	0.070	0.189	0.072	rs459920	0.013	0.111	0.111
rs5030240	0.084	0.130	0.088	rs6451722	0.253	0.039	0.032	rs17642714	0.048	0.101	0.065
rs182549	0.050	0.167	0.054	rs1042602	0.072	0.144	0.089	rs870347	0.041	0.046	0.115
rs722098	0.108	0.147	0.002	rs10108270	0.241	0.032	0.024	rs2166624	0.144	0.010	0.014
rs5997008	0.206	0.009	0.040	rs7495174	0.009	0.111	0.163	rs174570	0.086	0.006	0.060
rs917118	0.164	0.016	0.030	rs4778241	0.019	0.131	0.087	rs671	0.028	0.044	0.079
rs1886510	0.032	0.114	0.050	rs26722	0.022	0.081	0.123	rs7722456	0.037	0.002	0.068
rs10843344	0.120	0.020	0.036	rs735612	0.075	0.004	0.104	rs2024566	0.016	0.048	0.028
rs10141763	0.024	0.101	0.048	rs1126809	0.045	0.080	0.047	rs4411548	0.003	0.019	0.054

(continued)

Table 2
(continued)

Common AIMs	AFR	EUR	E ASN	2. Lao AIMs	AFR	EUR	E ASN	4. FROGkb AIMs	AFR	EUR	E ASN
rs1024116	0.011	0.082	0.072	rs1393350	0.042	0.078	0.046	rs1079597	0.008	0.020	0.044
rs2065160	0.031	0.092	0.034	rs12896399	0.135	0.027	0.002	rs2042762	0.007	0.018	0.010
rs7897550	0.072	0.010	0.027	rs4918842	0.026	0.032	0.083	rs12439433	0.007	0.015	0.008
rs1498444	0.077	0.029	0.000	rs1540771	0.080	0.053	0.005	rs12498138	0.022	0.002	0.001
rs2304925	0.016	0.041	0.010	rs12821256	0.019	0.040	0.024	rs4471745	0.000	0.003	0.005

Common forensic AIMs found in multiple sets are listed in the six top left boxes, representing: (A) all four sets; (B) sets 1 + 3 + 4; (C) 2 + 3 + 4; (D) 1 + 3; (E) 2 + 3; and (F) 3 + 4

co-ancestry component proportions can be exaggerated for those populations. An example of this effect is shown by a study of Bolivians [22] using the 46-plex AIM-Indels and a 446 SNP genomic ancestry panel developed by the LACE Consortium [23] illustrated in Figure 2. Analysis of Bolivian data with *Structure* shows the small AIM set consistently underestimates American co-ancestry proportions and overestimates European co-ancestry proportions compared to estimates made with 446 SNPs. Such results arise from the stronger differentiation of Europeans than Americans from the Indel panel, whereas the LACE panel provides much greater balance from very closely matched differentiation of Africans, Europeans, and Americans.

To identify the most powerful AIM-SNPs, we selected the most highly differentiated markers for population comparisons extended to the five main global population groups. Differentiation, indicated by how divergent the population groups are for the markers chosen, can be assessed by estimating one or all of the

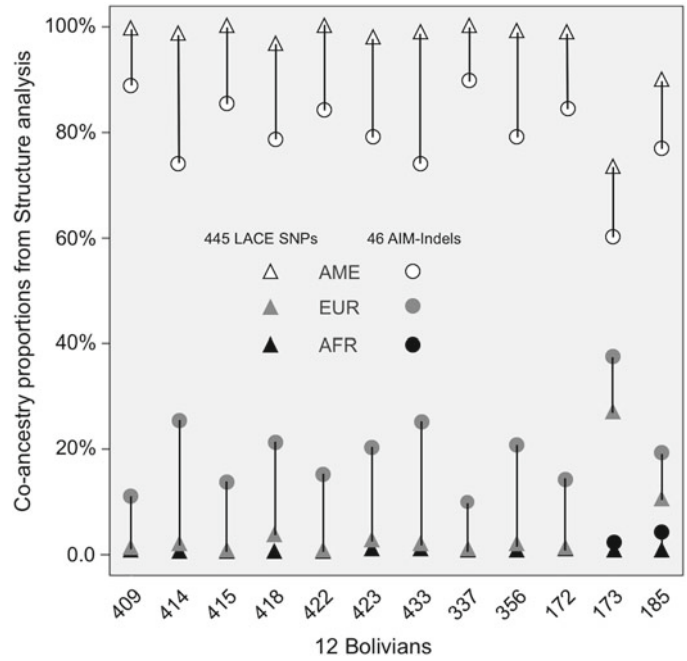


Fig. 2 Percentages of co-ancestry proportions from *Structure* analysis of 12 Bolivians inferred using different AIM-SNP sets: the 46 AIM-Indels and a larger scale genomics orientated 446-SNP panel developed by the LACE Consortium [23]. *Vertical bars* highlight the difference of co-ancestry proportions estimated with each marker set. In each individual the smaller-scale Indel set consistently underestimates American co-ancestry and overestimates European co-ancestry. Data indicates co-ancestry proportion estimation of admixed individuals with forensic scale AIM sets can be biased unless Population-Specific Divergence is carefully adjusted to achieve adequate balance. Figure adapted from Taboada-Echalar et al. [22]

following differentiation metrics for groups of candidate SNPs: Fst, delta (δ), (*see* **Note 4**) and Shannon's or Rosenberg's Divergence (the latter usually given as: *In*). The most informative markers have the highest Fst, δ, and *In* values for a particular population comparison and will best measure levels of admixture between those populations. Therefore we selected the SNPs showing highest differentiation metrics across all ten comparisons between the five global population groups. This was achieved by recording the cumulative Population-Specific Divergence values (PSD), i.e., applying to one group compared to all the others (*see* **Note 5**). Figure 3 compares the cumulative PSD values in five groups comparing 80 established forensic AIMs (34-plex and 80

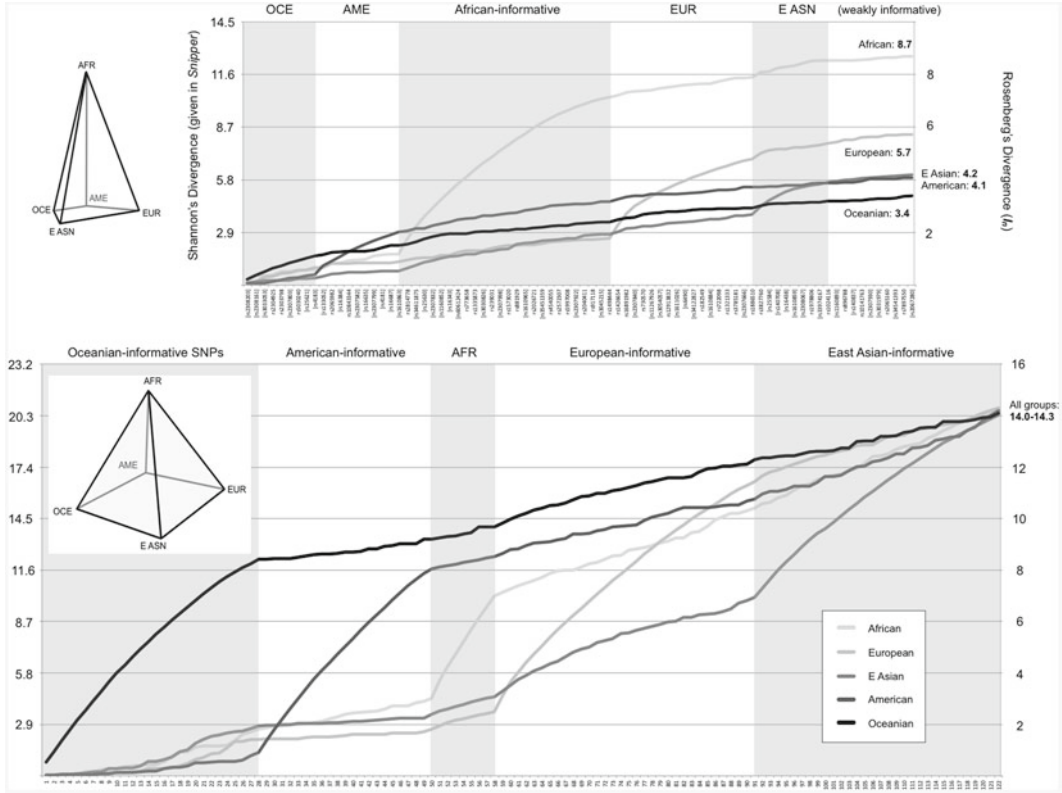


Fig. 3 Cumulative PSD values for 80 markers of 34-plex SNPs and 46 AIM-Indels (*upper chart*, Indels in *brackets*) and 122 loci of a "Global" AIM-SNP set developed for NGS. In each plot markers are ordered by informativeness to differentiate population groups: Oceanians; Americans; Africans; Europeans; East Asians, best to worst within each group. The pyramids *on the left* show geometric representations of the final cumulative PSDs achieved for each set—with 80 established AIMs showing strong skew in differentiation power towards African divergence, reduced amongst the others, particularly Oceanians. In contrast, 122 carefully selected AIMs can reach divergence balance (all pyramid points evenly distributed and PSDs near-identical). The final cumulative PSDs are given for each population group with the 122 AIM panel producing five values within 2 % of each other. Note that the slopes of the 122 AIMs PSD lines are straighter indicating comparable power amongst component loci

Indels combined) and 122 newly selected SNPs for NGS. The plots illustrate that established marker sets tend to accumulate African divergence more easily than American or Oceanian—generally SNP variation is less well differentiated in these two groups for the markers commonly selected as most powerful for Africans, Europeans, and East Asians (*see* Subheading 3.6). Additionally, when SNPs are targeted for differentiating American or Oceanian groups they tend to show greater divergence with Africans than with Europeans and East Asians. Therefore the proportions of SNPs differentiating each population group must be adjusted to reach a balanced final set with reduced tendency to overestimate co-ancestry from some populations.

To balance population differentiation it is possible to carefully maintain equilibrium for any one of the five groups by measuring the rising cumulative PSD as every new SNP is added. Generally PSD only rises for one population at a time and reaching a point of "convergence": where all values arrive at a point of near-identical cumulative PSD requires monitoring all population groups simultaneously and the accumulation of more SNPs for weakly differentiated groups such as Oceanians compared to more strongly differentiated groups such as Africans. Cumulative PSD values can be obtained in a very straightforward way by using the Divergence listing in *Snipper* for a full list of candidates then re-adjusting combinations until a point of convergence is obtained with a suitable number of loci (*see* **Note 5**).

3.7 Concluding Remarks

While SNaPshot has been the only viable forensic SNP typing system for over 12 years, the DNA analyst now has the opportunity to adapt recently developed NGS systems for sequencing and a variety of STR, SNP, and Indel genotyping applications. As NGS multiplexing levels are more than tenfold greater than those possible with the PCR and SBE reactions in SNaPshot, there is every opportunity to greatly improve data depth with the consequent enhancement of ancestry assignment likelihoods or analysis of individual admixture. Forensic practitioners interested in developing AIM sets will be able to identify and combine the best ancestry markers from existing forensic sets plus new forensic AIMs amongst the most powerful markers listed in SPS or published genomic ancestry sets such as the LACE panel [23].

To rebuild our own NGS forensic ancestry panel we focused on differentiating the five main population groups representing the most populace regions of the world, broadly separated by continental margins or geographic barriers that are consistently shown to match the global division of genetic variability [24]. This AIM set of 122–128 markers will require a separate additional panel of AIMs for the further subdivision of Eurasia into South Asia or Middle East distinct from Europeans, as marker selection did not set out to make this differentiation. Nevertheless supplementary

AIM-SNP sets such as *Eurasiaplex* go a long way to addressing this need already [8] and will form part of an additional set for subpopulation differentiation within Eurasia.

In parallel to the expanded SNP genotyping scope promised by NGS systems, *Snipper* has been enhanced to analyze larger numbers of SNPs with additional analysis options for more detailed study of population admixture when this is detectable in the individual SNP profiles submitted. These developments are described in detail in the second of the forensic ancestry analysis chapters (Chapter 19) that focuses on the statistical analysis of genetic data generated by the markers and multiplexes reviewed here.

4 Notes

1. Genetic analysis of ancestry provided in support of Operation Minstead—the 18-year investigation of a serial rapist in South London—provides a notable example of poor eyewitness reliability despite a large number of victims attacked. This was due to predominantly elderly victims lacking visual acuity in nearly all cases, plus a modus operandi that involved forced household entry at night, with the electricity cut and using a mask. Many victims also failed to give a reliable account of skin color, so likely pigmentation patterns also became a part of the DNA analysis provided from the extended SNP-based ancestry tests that included key coding SNPs: rs1426654 (SLC24A5); rs16891982 (SLC45A2); rs1042602 (TYR). Most forensic AIM-SNP sets include the first two SNPs since these show the highest differentiation of Europeans from other groups and to some extent provide an indication of likely skin color in admixed individuals that include European co-ancestry—although these type of predictions require much more thorough studies than currently exist to confirm why these loci give Europeans pale skin, but not East Asians.
2. At the time of writing 1000 Genomes is expected to release SNP data for new populations: CDX; KHV; GWD; MSL; ESN; ACB; PEL; GIH; PJL; BEB; STU; ITU (acronyms detailed at: http://www.1000genomes.org/category/frequently-asked-questions/population). SPS will upload this data as soon as it is released. Additionally, forensic Indel data is being prepared for dedicated SPS pages for scrutiny and construction of training sets for the AIM-Indel set described here. Regarding the modified 34-plex test described in [7], SNP genotype data has been kept available in both the SNP*for*ID SPSmart pages and in *Snipper* for the legacy set with rs727811 and the revised set with rs3827760. The "80-plex" option in *Snipper* combines the new 34-plex set with Indels.

3. Setting a probability threshold for an ancestry assignment made from Bayesian analysis of AIM genotypes is not an exact science. Likelihood ratios obtained are completely dependent on the training sets used in *Snipper*. For example, simple pairwise comparisons such as those made for the 11-M investigation where only two assignments were compared [2] will give higher probabilities than three, four, five, six, or seven population comparisons possible with HGDP-CEPG data (i.e., AFR-EUR-E ASN; +AME; +OCE; +South Asians; +Middle Eastern populations). Higher orders of magnitude differences can be used to reduce classification error than 100 times (two orders) but this obviously reduces the total number of assignments that are achieved applying small-scale AIM sets. Therefore a balance must be reached between lowest possible assignment error (too high if the threshold is set low) and the classification rate (too low if the threshold is set high). Assignment success rates from cross-validation of samples with known ancestries provide an indication of a practical threshold level that can be set for any one set of markers to minimize error and maximize the classification rate.
4. The simple delta allele frequency differential (δ) is useful as a quick assessment of candidate ancestry markers for a particular pairwise population comparison (δ = allele frequency in population 1 minus allele frequency in population 2). Delta correlates to Fst as $\mathrm{Fst} \approx \delta^2$ or $\mathrm{Fst} \approx \delta/(2-\delta)$. However, most studies use Rosenberg's *In* and this can be easily obtained by applying SNP genotype data from selected SPS population queries to *Snipper* cross-validation analysis (http://mathgene.usc.es/snipper/analysispopfile2_new.html) to obtain a ranked list of Divergence values. As these are Shannon's Divergences, they can be simply multiplied by 0.69 to get the equivalent *In* values.
5. Population-Specific Divergences (PSDs—also termed Locus-Specific Branch Length Divergences or LBSL) (*see* ref. 23) were obtained by creating training sets for *Snipper* from genotypes of candidate AIM-SNPs from the SPS ENGINES or HGDP-CEPH browsers (African, European, E Asian data available for all loci, American and Oceanian for 650,000 HGDP-CEPH SNPs). Training sets are labeled as African:non-African (i.e., all other groups), European:non-European, etc. accordingly, then saved as separate AFR, EUR, etc. Excel files. A cross-validation analysis of each file produces individual PSDs for each SNP and group comparison then cumulative values are obtained by their addition. Figure 3 shows the ranked PSD values for 80 markers (34-plex SNPs + 46 AIM-Indels) compared to those for 122 markers: forming the "Global" AIM-SNP set developed for NGS systems.

References

1. Phillips C, Salas A, Sanchez JJ, Fondevila M, Gómez-Tato A, Álvarez-Dios J, Calaza M, Casares de Cal M, Ballard D et al (2007) Inferring ancestral origin using a single multiplex assay of autosomal ancestry-informative marker SNPs. Forensic Sci Int Genet 1:273–280
2. Phillips C, Prieto L, Fondevila M, Salas A, Gómez-Tato A, Alvarez-Dios J, Alonso A, Blanco-Verea A, Brión M et al (2009) Ancestry analysis in the 11-M Madrid bomb attack investigation. PLoS One 4:e6583
3. Phillips C (2013) Ancestry informative markers. In: Siegel JA, Saukko PJ (eds) Encyclopedia of forensic sciences, 2nd edn. Academic, Waltham, pp 323–331
4. Rohlfs RV, Fullerton SM, Weir BS (2012) Familial identification: population structure and relationship distinguishability. PLoS Genet 8:e1002469
5. Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M (2010) IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye color in the absence of ancestry information. Forensic Sci Int Genet 5:170–180
6. Phillips C, Fondevila M, Lareu MV (2012) A 34-plex autosomal SNP single base extension assay for ancestry investigations. Methods Mol Biol 830:109–126
7. Fondevila M, Phillips C, Santos C, Freire Aradas A, Vallone PM, Butler JM, Lareu MV, Carracedo Á (2013) Revision of the SNP*for*ID 34-plex forensic ancestry test: assay enhancements, standard reference sample genotypes and extended population studies. Forensic Sci Int Genet 7:63–74
8. Phillips C, Freire Aradas A, Kriegel AK, Fondevila M, Bulbul O, Santos C, Rech Serrulla F, Perez Carceles MD, Carracedo Á et al (2013) *Eurasiaplex*: a forensic SNP assay for differentiating European and South Asian ancestries. Forensic Sci Int Genet 7:359–366
9. Santos NP, Ribeiro-Rodrigues EM, Ribeiro-dos-Santos AK, Pereira R, Gusmão L et al (2010) Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INDEL) ancestry-informative marker (AIM) panel. Hum Mutat 31:184–190
10. Pereira R, Phillips C, Pinto N, Santos C, Santos SEB, Amorim A, Carracedo Á, Gusmão L (2012) Straightforward inference of ancestry and admixture proportions in individuals and populations: a panel of 46 Insertion Deletion Ancestry Informative Markers typed in a single reaction. PLoS One 7:e29684
11. Phillips C, Fernandez-Formoso L, Gelabert-Besada M, Garcia-Magariños M, Santos C, Fondevila M, Carracedo Á, Lareu MV (2013) Development of a novel forensic STR multiplex for ancestry analysis and extended identity testing. Electrophoresis 34:1151–1162
12. Fondevila M, Phillips C, Santos C, Pereira R, Gusmão L, Carracedo Á, Butler JM, Lareu MV, Vallone PM (2012) Forensic performance of two Insertion-Deletion marker assays. Int J Legal Med 126:725–737
13. Zaumsegel D, Rothschild MA, Schneider PM (2013) A 21 marker insertion deletion polymorphism panel to study biogeographic ancestry. Forensic Sci Int Genet 7:305–312
14. Lowe AL, Urquhart A, Foreman LA, Evett IW (2001) Inferring ethnic origin by means of an STR profile. Forensic Sci Int 119:17–22
15. Pereira L, Alshamali F, Andreassen R, Ballard R, Chantratita W, Cho NS, Coudray C, Dugoujon J, Espinoza M et al (2011) PopAffiliator: online calculator for individual affiliation to a major population group based on 17 autosomal short tandem repeat genotype profile. Int J Legal Med 125:629–636
16. Phillips C, Fernandez-Formoso L, Garcia-Magariños M, Porras L, Tvedebrink T, Amigo J, Fondevila M, Gomez-Tato A, Alvarez-Dios J et al (2011) 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 5:155–169
17. Londin ER, Keller MA, Maista C, Smith G, Mamounas LA, Zhang R, Madore SJ, Gwinn K, Corriveau RA (2010) CoAIMs: a cost-effective panel of ancestry informative markers for determining continental origins. PLoS One 5:e13443
18. Rosenberg NA, Li LM, Ward R, Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. Am J Hum Genet 73:1402–1422
19. Lao O, Vallone PM, Coble MD, Diegoli TM, van Oven M, van der Gaag KJ, Pijpe J, de Knijff P, Kayser M (2010) Evaluating self-declared ancestry of U.S. Americans with autosomal, Y-chromosomal and mitochondrial DNA. Hum Mutat 31:E1875–E1893
20. Gettings KB, Lai R, Johnson JL, Peck MA, Hart JA, Gordish-Dressman H, Schanfield MS, Podini DS (2014) A 50-SNP assay for biogeographic ancestry and phenotype

prediction in the U.S. population. Forensic Sci Int Genet 8:101–108

21. Nievergelt CM, Maihofer AX, Shekhtman T, Libiger O, Wang X, Kidd KK, Kidd JR (2013) Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel. Investig Genet 4:13
22. Taboada-Echalar P, Álvarez-Iglesias V, Heinz T, Vidal-Bralo L, Gómez-Carballa A, Catelli L, Pardo-Seco J, Pastoriza A, Carracedo Á et al (2013) The genetic legacy of the pre-colonial period in contemporary Bolivians. PLoS One 8:e58980
23. Galanter JM, Fernandez-Lopez JC, Gignoux CR, Barnholtz-Sloan J, Fernandez-Rozadilla C, Via M, Hidalgo-Miranda A, Contreras AV, Uribe Figueroa L et al (2012) Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. PLoS Genet 8:e1002554
24. Phillips C, Parson W, Lundsberg B, Santos C, Freire-Aradas A, Torres M, Eduardoff M, Børsting C, Johansen P, et al (2014) Building a forensic ancestry panel from the ground up: The EUROFORGEN Global AIM-SNP set. Forensic Sci Int Genet 11:13–25

Chapter 19

Inference of Ancestry in Forensic Analysis II: Analysis of Genetic Data

Carla Santos, Chris Phillips, A. Gomez-Tato, J. Alvarez-Dios, Ángel Carracedo, and Maria Victoria Lareu

Abstract

Three approaches applicable to the analysis of forensic ancestry-informative marker data—*STRUCTURE*, principal component analysis, and the *Snipper* Bayesian classification system—are reviewed. Detailed step-by-step guidance is provided for adjusting parameter settings in *STRUCTURE* with particular regard to their effect when differentiating populations. Several enhancements to the *Snipper* online forensic classification portal are described, highlighting the added functionality they bring to particular aspects of ancestry-informative SNP analysis in a forensic context.

Key words Genetic ancestry, Reference data, SPSmart browser, Bayesian methods, *STRUCTURE*, *Snipper*, PCA

1 Introduction

Classifying individuals into populations is often useful in population genetics applications. But the definition of populations is commonly subjective, based on linguistic, cultural, or physical characters, as well as the geographical location of sampled individuals. This is a sensible way of incorporating diverse types of information but it may be difficult to know whether a given assignment of individuals to populations based on these subjective criteria matches an assignment in genetic terms. For this reason, it can be useful to confirm that the subjective classifications are consistent with genetic information and hence appropriate for the intended classification regime [1, 2]. A possible approach starts with a set of predefined populations and then classifies individuals

Electronic supplementary material: The online version of this chapter (doi:10.1007/978-1-4939-3597-0_19) contains supplementary material, which is available to authorized users.

William Goodwin (ed.), *Forensic DNA Typing Protocols,* Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_19,

of unknown origin into these populations. This involves sampling DNA from members of potential source populations to estimate allele frequencies in each population at a series of unlinked loci. Allele frequencies can be used to compute a set of likelihoods that a given profile of genotypes originates in each population. These likelihoods allow the assignment of individuals of unknown origin to populations based on the highest likelihood ratio [2].

Bayesian population analysis methods infer a simple relationship between the allele frequencies of a population and the allele frequencies observed in the individuals identified as part of that population. An advantage of such methods is that prior information about the samples can be used to progress the analysis. But the ability to differentiate populations in a sample set can be limited when applying a small number of samples and/or markers. Two valid approaches for comparing profiles from forensic casework DNA to reference population data will be considered here: a systematic Bayesian clustering approach (*STRUCTURE* software) and a naïve Bayesian likelihood ratio (LR) based calculator (underlying the *Snipper* web portal). *STRUCTURE* is a flexible approach—different types of markers such as STRs, SNPs, and indels can be readily combined in the same genotype input file (*Snipper* also allows such flexibility but the systems are not yet implemented). However *STRUCTURE* analysis of single profiles, typical of forensic testing, is not so straightforward since the whole set of parental data plus the unknown profile must be re-analyzed in combination each time and this can be both time-consuming and cumbersome to perform for a small number of casework samples in turn. For this reason *Snipper* (http://mathgene.usc.es/snipper/) was developed to provide a simple alternative for making ancestry assignments of single profiles in real time. Both *STRUCTURE* and *Snipper* use a Bayesian approach which, put simply, computes likelihood of membership to each class (in this case ancestry) using the observed frequency of variables in each class (in this case allele frequencies). The difference between both methods lies in how the likelihood is computed (more information about these algorithms is detailed in [2, 3]). Therefore both algorithms require reference data to calculate allele frequencies for comparison to alleles recorded in profiles of unknown origin. In the case of *Snipper*, the reference data allows construction of training sets for calculation of allele frequencies and these can comprise ready-to-use fixed five-population group data (African, European, East Asian, Native American, and Oceanian) already in place for 34 SNPs [3, 4] and/or 46 AIM-indels [5]. It can alternatively consist of end user's own data for any populations and binary marker set where reference genotypes are available, which can then be uploaded as a custom data set. Each algorithm makes the same prior assumption, often untested: that the variables, i.e., the component loci, are independent. For this reason, uniparental data (in

the form of haplotypes where all markers are linked) is not readily incorporated into either analysis system, though *STRUCTURE* has scope for the analysis of linked loci.

As the number of populations increases, the number of dimensions needed to visually represent the pairwise genetic distances also increases. The main idea of multivariate analyses is to help to represent, in a comprehensive way, those multiple dimensions. This is done through the reduction of the dimensionality of a data set composed of a large number of interrelated variables maintaining the maximum proportion of the variation present on that data [1]. Principal component analysis (PCA) is a commonly used multivariate analysis method, especially as an exploratory tool and to summarize genetic similarities and differences between groups of populations. This is possible through the transformation of those variables into a new set of metrics (principal components: PCs) that are not related and can be ordered in a way that the first PCs retain most of the variation present in the original data—the graphical representation of the first two or three PCs summarizes as much of the variation as possible in a comprehensive, graphical way [1, 6]. In the graphics that PCA generates, individuals are represented by points distributed according to their coordinates in two-way or three-way PC comparisons (two or three dimensional plots, respectively). PCA can also be used to represent the relation of an unknown study sample with a set of reference population samples, i.e., the study sample will be represented by a point superimposed onto the PCA plot of the reference population samples.

2 Materials

1. *SPSmart* browser home: http://spsmart.cesga.es/
2. *SPSmart* SNP*for*ID 52-plex and 34-plex variability browser: http://spsmart.cesga.es/snpforid.php
3. Entire genome interface for exploring SNPs (ENGINES) a 1000 Genomes variability browser: http://spsmart.cesga.es/engines.php?dataSet=engines
4. pop.STR: http://spsmart.cesga.es/popstr.php
5. *Snipper* portal: http://mathgene.usc.es/snipper/
6. *STRUCTURE* software: http://pritchardlab.stanford.edu/structure.html
7. *Structure harvester*: http://taylor0.biology.ucla.edu/structureHarvester/#
8. CLUster Matching and Permutation Program (*CLUMPP* software): http://www.stanford.edu/group/rosenberglab/clumpp.html

9. *distruct* software: http://www.stanford.edu/group/rosenberglab/distruct.html
10. For more information about *R* software: http://www.r-project.org/ [7].

3 Methods

3.1 Collection of Ancestry Reference Data with the SPSmart Browser

The statistical analysis of a profile requires reference training sets, i.e., parental populations of interest used to classify casework profiles. Collection of such data previously required locus-by-locus scrutiny of dbSNP or HapMap SNP databases [8], but fortunately SPS (SNPs for Population Studies) makes this task much more straightforward for any number of AIM-SNPs as well as up to 52 STRs in routine forensic use.

SPS comprises the following genomic variability browsers:

- SPSmart home: http://spsmart.cesga.es/ [9].
- SPSmart SNP*for*ID 52-plex and 34-plex variability browser: http://spsmart.cesga.es/snpforid.php [10].
- ENGINES (Entire Genome Interface for Exploring SNPs) a 1000 Genomes variability browser enabling a review of all SNP sites found from 1092 complete genome sequences (1000 Genomes Phase I): http://spsmart.cesga.es/engines.php?dataSet=engines [11].
- pop.STR: http://spsmart.cesga.es/popstr.php [12].

SPSmart [9] is a simple pre-processing engine that includes five different population-based genotype databases: (1) 1000 Genomes Phase I May 2011; (2) HapMap Release #28; (3) Perlegen complete data set; (4) HGDP-CEPH Stanford study; and (5) HGDP-CEPH NIH-Michigan study (*see* **Note 1**). *SPSmart* also generates common population genetics indices such as allele frequencies, heterozygosity, F_{ST}, or *In* (summarized in the downloads tab of each query).

1. Choose the database(s) to search or choose SEARCH in the SNP*for*ID "global map" homepage.
2. Choose the populations to merge into groups by ticking selections up to a maximum of five. If opting to review multiple databases, e.g., HapMap and Perlegen, only one population grouping can be made. Populations are already arranged into sets of genetic diversity based on Rosenberg's original analyses of HGDP-CEPH populations [3, 13], though note that Eurasians are subdivided into European, South Asian, and Middle Eastern subgroups.
3. Add the SNP RefSeq (rs-number) identifiers in the search by SNP window. Search by chromosome region or gene name is also available. Click the "next" button below.

4. Of the filter options presented, MAF is minimum allele frequency and *In* Rosenberg's ancestry informativeness metric [14]. When reviewing SNP data from multiple databases, it is best not to tick option: "Filter SNPs not genotyped on every compared dataset".
5. When SNPs are not found from a query, a *message tab* with the missing rs-numbers appears.
6. In the *downloads* tab, genotype files are available ready to download, copy and paste into Excel (*see* **Note 2**) or notepad. The recommended steps being: download, choose all, copy into Excel, transpose the data into rows = samples and columns = SNPs (in edit menu: "copy" > select new cell > "paste special" > select: "transpose"). This must be completed for each group while taking care to label each set of sample rows with the appropriate description, e.g., African, South Asian, etc. Another option is to query and download all population groups at once—follow the steps previously described in this point for the genotypes and then download the sample list file. This file includes sample, subpopulation, and population group information. Open, select all, and paste in the genotypes Excel file—confirm that the samples are in the same order and remove the duplicated column with sample name (Fig. 1).

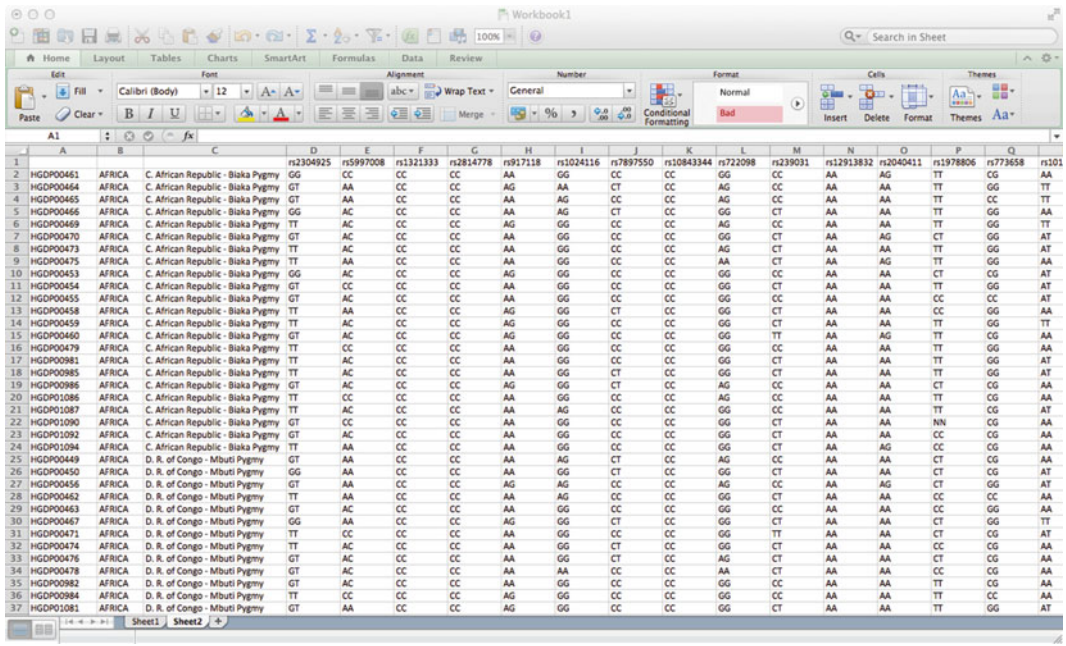


Fig. 1 Example of a reference ancestry genotype data file obtained from *SPSmart*. The data obtained from *SPSmart* was reorganized (original data was transposed so that samples are now organized in rows and markers in columns) and population information (downloaded from *SPSmart* in a separate file) was added

7. The genotypes can be formatted for input to *STRUCTURE*, *Snipper*, or PCA custom data analysis (outlined later). SNaPshot genotypes may need checking against the reference data if they come from different typing platforms, e.g., a SNaPshot C/T SNP may be an A/G SNP in HapMap requiring base inversions of one dataset. For this reason symmetrical SNPs (C/G and A/T) require particular care (*see* **Note 3**).

3.2 STRUCTURE Software

3.2.1 Background on STRUCTURE Analysis

STRUCTURE uses genotypic data of several loci to: (1) infer population structure; (2) identify subpopulations (subsets of samples with distinct allele frequencies); (3) assign individuals to subpopulations (based on probabilities); and (4) study admixture between populations. It uses a population structure model where studied samples represent a mixture of K unknown populations—each characterized by unknown allele frequencies for the loci used and where these are assumed to be in Hardy-Weinberg equilibrium (HWE) and independent (not in linkage disequilibrium). The objective is to classify individuals into K clusters in a way that deviations from HWE and independence are minimized.

Assuming HWE and independence in each subpopulation, the probability that the genotype of an individual belongs to subpopulation k is given by the product of the allele frequencies. Using Bayes rule (*see* **Note 4**) it is possible to calculate the probability of an individual belonging to subpopulation k. If allele frequencies in a population were known in advance, it would be easy to allocate individuals. Equally, if individual allocations were known it would be easy to estimate the frequencies. In practice, we do not know either, but using a Markov Chain Monte Carlo (MCMC) algorithm (*see* **Note 5**) it is possible to obtain a sensible estimate of both.

The estimation of the optimum K value is performed in an independent way: for each simulation a posterior probability value $Pr(K)$ is calculated. In general, for K values under the optimum, $Pr(K)$ is low but it tends to stabilize with higher K values, so a *plateau* is commonly observed. When several K values have similar $Ln\ Pr(K)$ estimates, the smallest of them is usually the most appropriate estimate—generally corresponding to the inflection point of the probability plot. However, it is not always possible to know the real K value, so it is best to choose the lowest value that captures the maximum structure present in the data [15]. Evanno et al. described a method to estimate K based on a second-order change of probability value taking into account the variability of the probability value between different replicates of each K value—*delta K* [16]. This method is implemented in *Structure Harvester* [17] (*see* **Note 6**). However K is not an absolute value, defined values should be carefully considered taking into account any known characteristic of the studied populations. There are several factors that can affect the clustering of the samples: (1) number of markers; (2) number of samples; (3) number of clusters; and (4) allele frequency

correlation. Endogamy or genotyping errors can have the same effects as true population structure as they can simulate linkage disequilibrium in unlinked markers and deviations from HWE.

STRUCTURE has several models for ancestry (Fig. 2, *see* **Note 7**) and allele frequencies (*see* **Note 8**). For more details refer to the *STRUCTURE* manual and articles describing the different models [2, 18–20]. Considering the type of analysis required in a forensic context, the *admixture POPFLAG* ancestry model is appropriate. This combines two important features: consideration of admixture between populations (individuals can have recent ancestors from multiple populations so ancestry membership proportions from each ancestral population can be calculated); and some individuals can be used as a reference to help infer the ancestry of the samples under study. Regarding the allele frequencies model, it is advisable to use the correlated allele frequencies model because it will guarantee that an undetected correlation will be identified without affecting the results should it be absent.

3.2.2 Preparation of a STRUCTURE Input File

Data to be analyzed with *STRUCTURE* needs to be organized in a single matrix (as a text file) where optional information can be considered to complement the genotypic data. Such information should be included in a predefined order and it is important to highlight that only the genotypic data is required for the analysis. We will focus the construction of an input file on the information of greater relevance when analyzing a casework profile (Tables 1 and S1). For more information about constructing *STRUCTURE* input files (especially formatting information on recessive alleles, marker distance, phase information, or phenotype), refer to *STRUCTURE* software manual or to a recent overview [20].

- First line: header line. Headers are only included in the markers columns.
- First column: sample name information that can be an alpha-numerical code which can introduce errors when

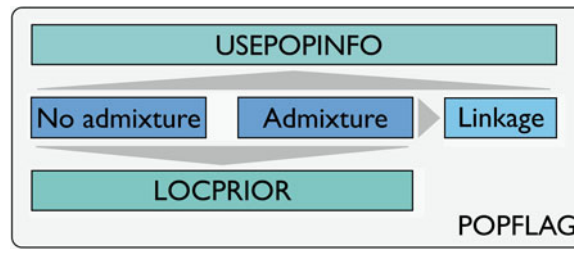


Fig. 2 Schematic representation of *STRUCTURE* ancestry models and their relationship. The central models are *no admixture* and *admixture*; both can be used together with *LOCPRIOR* information. The *admixture* model is the basis for the *linkage* model. All three models (*no admixture*, *admixture*, and *linkage*) can be used in conjunction with the *USEPOPINFO* model. All the above models can be used considering *POPFLAG* information

Table 1
***STRUCTURE* input file format**

						M1	M2	M*n*
S1	1	1	1	Extra	Extra	1	2	1
S1	1	1	1	Extra	Extra	3	4	1
S2	1	1	2	Extra	Extra	3	2	2
S2	1	1	2	Extra	Extra	3	2	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
S10	1	1	3	Extra	Extra	1	4	1
S10	1	1	3	Extra	Extra	1	4	2
S11	2	1	4	Extra	Extra	1	2	2
S11	2	1	4	Extra	Extra	3	4	2
S12	2	1	5	Extra	Extra	3	2	1
S12	2	1	5	Extra	Extra	3	2	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
S20	2	1	6	Extra	Extra	1	4	2
S20	2	1	6	Extra	Extra	1	4	2
S21	*n*	0	7	Extra	Extra	1	2	1
S21	*n*	0	7	Extra	Extra	2	4	1
S22	*n*	0	8	Extra	Extra	2	2	2
S22	*n*	0	8	Extra	Extra	2	2	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
S*n*	*n*	0	9	Extra	Extra	1	4	**-9**
S*n*	*n*	0	9	Extra	Extra	1	4	**-9**

Samples S1…S*n* from populations 1…*n* analyzed with genotypic data from markers M1…M*n*. Samples 1…10 belong to population 1 and can be divided into three locations (1–3). Samples 11…20 belong to population 2 and can be divided into three locations (4–6). Samples 21…S*n* belong to population *n* and can be divided into three locations (7–9). Samples from populations 1 and 2 are reference (*POPFLAG* = 1) and study samples are from population *n* (*POPFLAG* = 0)

running *CLUMPP*, easily solved in *Structure Harvester* (*see* **Note 6**).

- Second column: a numerical code representing the population of origin as defined by the researcher. By default this information is not used by the clustering algorithm but can help organize the output file.
- Third column: *PopFlag* information. This is a Boolean variable where 1 (TRUE) represents the samples that should be used as reference and 0 (FALSE) the casework/study samples.

- Fourth column: *LocPrior* information. A numerical code that denotes subpopulation groups, geographical locations, or other shared characteristic between individuals inside the population groups defined in the second column. This information is used when considering the *LOCPRIOR* ancestry model.
- Any number of extra columns can list useful information for the researcher. For example, as the population and *LocPrior* information is numeric, extra columns with the names can be included as an easy way to cross check data later.
- The following columns include genotypic data for any number of markers (SNPs, indels, or multiallelic markers such as STRs). Genotypes should be coded as numbers. For SNPs, we routinely use *A = 1*, *C = 2*, *G = 3*, and *T = 4*. STRs are already numerically coded but in the case of intermediate alleles the "." should be removed, i.e., *19.3 = 193* (*see* **Note 9**).
- Missing data is usually coded as *-9* but any other code not present in the file can be used.
- Each allele needs to be represented in a separate cell: both alleles in the same line but in different columns or both alleles in the same column but in different lines (we will focus on the latter as shown in Table 1).
- Spaces should not be included.

3.2.3 How to Run STRUCTURE Software

The first stage when running *STRUCTURE* (http://pritchardlab.stanford.edu/structure.html) is to create a new project (File > New project) following four established steps:

- Step 1—project information: name of the project, directory where the project will be saved, and input file.
- Step 2—information of input data set (*see* **Note 10**): number of individuals, ploidy of data, number of markers, and missing data value.
- Steps 3 and 4—format of input data set (*see* **Note 11**): information contained in rows and columns (e.g., row of marker names or individual ID for each individual).

Before creating the project, *STRUCTURE* presents a summary where it is possible to confirm the selected options. If there are no errors, the project opens and the data is visible.

The next stage is to create a new parameter set (Parameter set > New):

- Run length—a *burnin* period of 100,000 is more than sufficient to allow a progressive convergence towards reliable allele frequency estimates in each population and probabilities for membership of individuals to a population. Measurement of the assumed number of populations uses the MCMC estimation and is performed separately from

the *burnin*. About 100,000 MCMC repeats have been shown to provide good ancestry membership proportions estimates. But *burnin* and MCMC repeat number should be adjusted depending on the study objectives and information contained in the data set (*see* **Note 12**).

- Select ancestry model—depending on the study objectives and the data to be analyzed, different ancestry models can be considered. For the *admixture POPFLAG* model select "Use admixture model" under the "Ancestry model" tab and "Update allele frequencies using only individuals with *POPFLAG* = 1 data" under the "Advanced" tab.
- Select allele frequencies model—for the *correlated allele frequencies* model select the "Allele frequencies correlated" option under the "Allele frequencies model" tab.
- Leave the "Compute probability of the data (for estimating *K*)" option under the "Advanced" tab selected so that posterior *Ln Pr*(*K*) values are calculated—those will be used to estimate the optimum *K* value.
- Save the new parameter set with the desired name and confirm the selected options in the summary window that opens after saving. A tree on the left side of the screen will include all the parameter sets created, indicating the one active at the moment.

There are two ways of starting a simulation:

- Run a single *K* value—in the "Parameter set" menu select "Run" and set the assumed number of populations (*K*). This option only allows a single *K* value and replicate at a time.
- Schedule multiple runs—in the "Project" menu select "Start a job". A new window opens—select the parameter set(s) to be analyzed, the *K* values, and the number of iterations for each *K* (*see* **Note 13**). For example, two different parameter sets can be programmed to run from *K* = *2* to *K* = *6*, three replicates for each *K*—this sums up to 30 scheduled runs. This option is advantageous for large projects—a new run starts automatically after the previous one has finished so there is no need for constant attention on the progress of the job (*see* **Note 14**).

3.2.4 STRUCTURE Associated Software

Software associated with *STRUCTURE*, for example, *CLUMPP* [21], contains three algorithms for the alignment of multiple replicate analyses of the same data set which allows the transformation of any number of replicate simulations for each *K* in a single set of data (*see* **Note 15**). Such data is suitable for analysis with another supporting program *distruct* [22] which allows the visualization of the estimated membership coefficients: populations are represented as colors and individuals as bars portioned into colored segments that correspond to membership coefficients in the groups (*see* **Note 16**).

3.2.5 What Information Can Be Obtained from STRUCTURE?

STRUCTURE output files include information on the estimated clusters, i.e., the population groups generated not the input populations. However when the populations are defined in a way that they closely match the calculated clusters, the inferences of the population ancestry membership proportions in each of the predefined clusters can be considered to be the proportions of the input populations. When attempting to classify a population or individual, the use of reference populations closely matching the inferred clusters is important, especially when analyzing admixed samples where it is important to define the contributing parental populations. The ancestry membership proportions for each individual in each cluster are also calculated by *STRUCTURE*.

Allele frequency divergence among populations, average distances (expected heterozygosity) between individuals in the same cluster, mean F_{ST} values, and estimated allele frequencies in each cluster (including estimated ancestral frequencies) are calculated. As a way of quantifying the information given by a particular *STRUCTURE* run and estimating the optimum K value it calculates the estimated probability of the data, the mean likelihood value, and associated variance. And it calculates the mean value of *alpha* (α) as a measure of the relative admixture levels between populations—when $\alpha >> 1$ the individuals are highly admixed; for values of $\alpha << 1$ each individual has its origin mainly in one population (from our experience with the HGDP-CEPH panel of samples, $\alpha < 0.05$—this value varies depending on the population groups considered and the differentiation power of the marker sets used).

The population and individual ancestry membership proportions can be represented in two distinct types of plot:

- A bar plot where each individual of the data set is represented by a vertical line divided into K colored segments proportional to the estimated membership into each of the K inferred clusters. To visualize the bar plot in *STRUCTURE* choose the appropriate result file in the tree on the left side of the window—on the simulation result window menu select Bar plot > Show.
- Each individual is represented as a colored point in a triangle (on the simulation result window menu select Triangle plot > Show). Colors correspond to the population tag in the input file. The estimated ancestry vector for an individual is formed by K components that sum up to 1. This type of plot is particularly useful to represent $K = 3$ data because the vectors can be represented in one triangular plot. For each point, the distance to the triangle vertices gives each of the three components. Individuals located in one of the vertices are completely assigned to the population represented in it.

Despite the advantages of the triangular plot when visualizing $K = 3$ data, bar plots are usually easier to interpret, especially for $K > 3$.

In the case of forensic casework analysis, *STRUCTURE* gives information on the training set (allowing the assessment of the used reference data set—the optimum K value matches the number of reference populations, which are completely differentiated among them) and it also gives us the individual ancestry membership proportions (such information has considerable potential in guiding investigators to more clearly defined suspect pools, this being particularly true when no eyewitness is available or STR profiles fail to match DNA database records). This is illustrated in Subheading 3.5.

3.3 The Snipper Web Portal

3.3.1 Background on Snipper Analysis

The *Snipper* portal includes a straightforward Bayesian system for predicting ancestral origin and estimating the misclassification rate. It uses a set of samples of each population as training sets and assigns individuals to the population that maximizes the posterior probability (maximum likelihood calculation) [3]. The likelihood parameters are estimated from training set allele frequencies assuming HWE and independence for the used loci (*see* **Note 17**).

Snipper was originally designed to provide a real-time ancestry assignment system for 34-plex profiles with reference to default pre-typed AFR-EUR-E ASN training sets and this still represents the simplest approach for assessment of a single casework profile to obtain an immediate overview of ancestry. The portal has been updated to include 34-plex [3, 4] and AIM-indel [5] fixed reference data for five populations groups: AFR-EUR-E ASN-AME-OCE. But the ancestry analyses can be extended beyond the default settings. For example, custom Excel files (including any binary markers that are of interest for the researcher) or frequency based Excel files (helpful when working with STRs or haplotypes) can be used as reference training sets.

A new version of *Snipper* is being prepared (*Snipper App suite version 2.0*) to include new functionalities including turn on/off the HWE assumption; prediction of admixture components; batch analysis (multiple profiles); fine-tuning of a training set; classification of single profiles; and analysis of training sets through multinomial logistic regression (beta version). At the time of writing a publication describing *Snipper 2.0* is in preparation.

3.3.2 Preparation of a Snipper Input File

Careful preparation of the Excel file containing the custom training set profiles and precise matching of unknown profiles to training set data for bases and locus order is important. Therefore it is recommended to sort component SNPs/indels into ascending rs-number order as an aid to data checking.

For *Snipper* analysis using binary markers, an *.xlsx* Excel file (*.xls* can still be used for certain previous options) with sample, population, and genotype information listed (Tables 2 and S2)—how that information is organized is also important so the following considerations should be taken into account:

- Cell 1A indicates the number of samples; cell 1B the number of markers; and cell 1C the number of populations.
- Line 1 (from column D onwards) specifies the marker name (represented by an alpha-numerical code).
- Lines 2–5 can be left empty or can be used to include useful notes (e.g., one of the lines can be used to store the study/casework sample profile and other line can be used

Table 2
***Snipper* input file format**

	A	B	C	D	E	…	XFC	XFD
1	# Samples	# Markers	# Populations	M1	M2	…	M*n*	
2								
3			Profile	AG	TT	…	AC	
4			Concatenate	=D3&E3&…&XFC3				
5								
6	1	P1	S1	AG	CT	…	AA	1
7	2	P1	S2	GG	CC	…	CG	1
8	3	⋮	⋮	⋮	⋮	…	⋮	⋮
9	4	P1	S10	AA	TT	…	AG	1
10	5	P2	S11	AG	CT	…	GG	1
11	6	P2	S12	GG	CC	…	AC	1
12	7	⋮	⋮	⋮	⋮	…	⋮	⋮
13	8	P2	S20	AA	TT	…	CC	1
14	9	P*n*	S21	AG	CT	…	AA	0
15	10	P*n*	S22	GG	CC	…	GG	0
⋮	⋮	⋮	⋮	⋮	⋮	…	⋮	⋮
1048576	*n*	P*n*	S*n*	AA	TT	…	**NN**	0

Samples S1…S*n* from populations 1…P*n* analyzed with genotypic data from markers M1…M*n*. Samples 1…10 belong to population 1; samples 11…20 belong to population 2, and samples 21…S*n* belong to population P*n*. An extra column after the last marker (in this case column XFD) should be included when trying to classify several study samples simultaneously—samples from populations 1 and 2 are reference (labeled as 1) and samples from population P*n* are the unknown study (labeled with 0). Lines 2–5 can be used to include useful information—e.g., when a single profile is being classified it can be included (here in line 3) and concatenated (cell D4)—the concatenated profile can then be copy-pasted directly into *Snipper*

to concatenate that profile—ready for copying and pasting, i.e., if the profile is in line 3 type = *D3&E3&F3…* in the desired cell).

- Column A (from line 6 onwards) has a numeric value that usually represents a sample.
- Column B (from line 6 onwards) has the population names.
- Column C (from line 6 onwards) has the sample names (which can be represented by an alpha-numerical code).
- Column D onwards (from line 6 onwards) includes the genotypes (coded as nucleotide bases—ACGT). Missing data should be coded as NN. Other symbols in the file (e.g. ?, spaces) are not recognized. Triallelic markers can be included in the analysis.
- A new batch analysis option was implemented in *Snipper v2.0* which allows for simultaneous classification of more than one profile. In this case, the input file should be constructed as described in the previous points. An extra column after the last marker (with no headers—start in line 6) needs to be included: training samples are to be marked as 1 and study samples to be classified as 0.

3.3.3 How to Run Snipper

Snipper includes several options to classify individuals and analyze populations. For forensic analysis the two most applicable options are: "*Classification as Europe-East Asia-Africa-America-Oceania (34 SNPs, 46 Indels, or both sets)*" and "*Classification with a custom Excel file of populations*". There is an additional option that works in the same way but allows batch analysis: "*Classification of multiple profiles with a custom Excel file of populations*".

1. The "*Classification as Europe-East Asia-Africa-America-Oceania (34 SNPs, 46 Indels, or both sets)*" option uses fixed training sets and provides a simple system to classify single profiles.
   - Step 1—go to http://mathgene.usc.es/snipper/pop-choosing5groups.html
   - Step 2—choose the marker set from three options: 34-plex SNPs (the original marker set [3] or the revised set [4] can be selected), 46-plex AIM-indels [5], or a combination of 80 binary markers (Indels combined with the revised 34-plex set). SNPs are listed in rs-number order and AIM-indels in electrophoretic order—on the left side links give images listing the marker order in each option.
   - Step 3—choose populations. Three to five main population groups are available (Africa, Europe, East Asia plus America plus Oceania).
   - Step 4—choose the classifier. Four options are now available: naïve Bayesian analysis (considering whether the Hardy-Weinberg principle applies or not), multinomial logistic regression, and genetic distance algorithm.

- Step 5—data input. Depending on the option selected in Step 1, a profile including 34, 46, or 80 markers (68, 92, or 160 bases respectively) should be typed (*see* **Note 18**). As described before, a profile can be built by concatenating data in Excel (using the "&" operand) allowing individual scrutiny of composite genotypes before direct copy-pasting into the query window left of the "*Classify*" button (*see* **Note 19**).

2. "*Classification with a custom Excel file of populations*"—this option allows extension of ancestry analyses beyond the default five-population group comparisons and 34, 46, or 80 binary markers using *Snipper*.
   - Step 1—go to http://mathgene.usc.es/snipper/analysispopfile_new.html
   - Step 2—data input (population). An Excel file prepared as described above (Table 2 without the final column) is uploaded.
   - Step 3—choose classifier. Options as described above.
   - Step 4—data input (individual). A profile string containing the same number of markers in the same order as they appear in the data file uploaded in Step 1 (*see* **Note 18**) is entered in the query window.
3. "*Classification of multiple profiles with a custom Excel file of populations*"—go to http://mathgene.usc.es/snipper/analysismultipleprofiles.html. This option works as above but without the need for individual profile submission. Profiles to be classified are indicated as previously described (Table 2). The multinomial logistic regression classifier function is not currently available for this option.

Snipper also includes an option to analyze training sets to gauge characteristics of the component binary markers—"*Thorough analysis of population data of a custom Excel file*" (http://mathgene.usc.es/snipper/analysispopfile2_new.html). This is useful to assess the informativeness of new candidate AIM binary markers for ancestry inference. After uploading the Excel file of custom data and defining Hardy-Weinberg, choose "*Perform a verbose cross-validation analysis of my population data with the best _ SNPs*" adding the relevant number of markers to assess. Cross-validation removes each component sample in turn, recalculates the allele frequencies in the training set, and then assigns ancestry for the removed profile. The other options "*Try to classify all individuals in the sample*", "*Perform a non verbose cross-validation analysis of my population data*", and "*Compute bootstrap error of my population data*" provide choice of alternative assignment error estimators. Multinomial logistic regression can also be applied to the population data—in this case information given in Step 2 about HWE will be ignored. Once a training set has been assessed for informativeness, users can choose options

2 and 3 described above to compare single or multiple profiles from unknown samples to the custom reference data and assign ancestry in identical fashion to using the fixed training sets.

3.3.4 Evaluating Snipper Output

Results from the analysis of a profile comprise the submitted profile; the assumed classifier; the -log likelihoods (use of -log likelihoods permits easier comparison of the very small likelihood ratio figures normally generated) and percentiles for the training set population groups; the likelihood ratios in verbose format and predicted admixture components and ancestry; a set of plots summarizing the classification; the apparent success of the classification; and a list of the markers in descending order of divergence (*see* **Note 20**). Missing genotypes are flagged in red in the divergence list to allow some assessment of the potential contribution of gaps in the profile, in other words, assignments made with several red markers at the top of the list will be much less reliable than those with gaps at the bottom, although this will be clear from the probabilities obtained. Apparent success measures the rate of correct assignment of training set samples using the markers of the profile. These values are 100 % for a complete set of markers, but drop when significant numbers of gaps occur in the submitted profile (in the case of the three group 34-plex fixed training set this is particularly true for EUR:E ASN comparisons).

3.4 Principal Component Analysis (PCA)

3.4.1 Background on PCA

Principal component analysis or PCA is a multivariate data analysis technique allowing the reduction of dimensionality, i.e., it uses fewer variables, while preserving much of information in the data. Usually two or three principal components are made, constructed as linear combinations of the original variables. Working with only two or three variables allows graphical representation of the data in a 2D plane or 3D graphic, providing fast visual recognition of patterns or clusters. Numerous software packages are available to perform PCA analysis when numerical variables are used. When SNP data is considered, an initial transformation (or recodification) is needed to access this existing software. The next section details SNP data preparation using the statistical package *R*.

3.4.2 Preparation of PCA Input Files

SNP analysis with PCA requires two text files with sample, population, and genotype information. One of the files should include training set data and the other the study samples to be compared. Both files have the same format (Tables 3, S3 and S4)—the system for organizing this data is important so the following considerations should be carefully taken into account:

- The first column includes sample name information in the form of an alpha-numerical code. The column header is "Sample".
- The second column has the populations/groups names. The header is "Population".

Table 3
Principal component analysis SNP input file format

Sample	Population	M_1	M_2	M*n*
S1	P1	AG	CT	AA
S2	P1	GG	CC	CG
⋮	⋮	⋮	⋮	⋮
S10	P1	AA	TT	AG
S11	P2	AG	CT	GG
S12	P2	GG	CC	AC
⋮	⋮	⋮	⋮	⋮
S20	P2	AA	TT	CC
S21	P*n*	AG	CT	AA
S22	P*n*	GG	CC	GG
⋮	⋮	⋮	⋮	⋮
S*n*	P*n*	AA	TT	**NN**

Samples S1…S*n* from populations P1…P*n* analyzed with genotypic data from markers M1…M*n*

- The following columns have genotype data, one marker per column. Each column header will have the corresponding marker name, which can be an alpha-numerical code. Genotypes are coded with nucleotide bases (ACGT) and missing data as NN. Note that markers must be in the same order in both input files.
- Spaces can be included in the file except as part of the genotype data (they will be considered as a new genotype, i.e., TT ≠ T T).
- Triallelic markers can be included in the input file but they will not be considered for the principal components calculation.

3.4.3 Creating a PCA Plot

In this subheading we include *R* scripts that can be used to generate 2D principal components graphics (only if the number of variables (SNPs) is smaller than the number of samples). The script commands can be copied and pasted into the *R* console. With the main focus on graphics, *R* offers a range of options to manipulate data and generate plots that adjust to user needs. In the case of this script, command lines were added to allow changes in color, shape, and size of the symbols representing individuals (*see* **Note 21**).

All text after the # symbol represents notes for the user and will not be computed. A *R* version of the script is included as supplementary in the digital version of this chapter.

```
##############################
### Script to make a 2D PCA ###
##############################

# Important Note: this script can only be used when the number of samples is equal to, or higher than, the
 number of SNPs.


# First open SNPassoc library - this is an association package that allows you to recode SNP data

library(SNPassoc)

TEST<-function(x){try(snp(x,sep=""),silent=TRUE)} # homemade function that detects if a SNP is bi- or tri-
 allelic

##############################
# To read the data from two input files. The computer will prompt for the location of the reference samples
 file and the study samples file, in that order
##############################

RefData<-read.table(file.choose(), sep="\t",header=TRUE,na.strings=c("NN")) # function that reads the
 reference data input file and stores it in the dat object

NRefSamples<-dim(RefData)[1] # get the number of reference samples

StudyData<-read.table(file.choose(), sep="\t",header=TRUE,na.strings=c("NN")) # function that reads the file
 with the study samples and stores it in the StudyData object

NStudySamples<-dim(StudyData)[1] # get the number of study samples

### Because tri-allelics are not considered when making the PCA, they are removed from both data sets using
 the function TEST.

apply(RefData,2,TEST)->RT

deletedSNPs<-which(as.numeric(summary(RT)[,1])==1) # defines which SNPs have to be removed

RefData2<-RefData[,-deletedSNPs] # for the reference samples, only the columns with bi-allelic SNP data are
 kept

StudyData2<-StudyData[,-deletedSNPs] # for the study samples, only the columns with bi-allelic SNP data are
 kept
# this removes tri-allelic SNPs if there were any

##############################
# To recode and typify the data
##############################

ComData<-rbind(RefData2,StudyData2) # combines the reference and study samples in one variable (ComData) -
 this needs to be done prior recoding the SNP data to guarantee the coding uniformity

datSNPT<-apply(ComData,2,function(x) {additive(snp(x,sep=""))}) # the additive function recodes each bi-
 allelic SNP in the data as numeric (0=homozygous for the most frequent allele, 1=heterozygous, 2=
 homozygous for the least frequent allele)

datSNP<-datSNPT[1:NRefSamples,] # after recoding, the reference data set is temporarily isolated into a new
 variable (datSNP) to make some computations

m<-apply(datSNP,2,mean,na.rm=TRUE) # this calculates a vector with the mean value of each "numeric" SNP

s<-apply(datSNP,2,sum,na.rm=TRUE) # this calculates a vector with the number of occurrences of the least
 frequent allele for each SNP

n<-apply(datSNP,2,function(x) sum(!is.na(x))) # this calculates a vector with the number of valid genotypes
 for each SNP (NN genotypes are not considered valid)

p<-s/(2*n) # vector with the frequency of least frequent allele for each SNP

XT<-scale(datSNPT,center=m,scale=sqrt(p*(1-p))) # this typifies the SNPs (to each "numeric" SNP value the
 average is subtracted and then divides by the standard deviation)
```

```
XT[is.na(XT)]<-0 # replaces missing values with 0

X<-XT[1:NRefSamples,] # after recoding and typifying, the reference data set is isolated into a new variable
 (X)

Y<-XT[-c(1:NRefSamples),] # after recoding and typifying, the study data set is isolated into a new variable
 (Y)

if(NStudySamples==1){ # this forces Y to be a matrix if there is only one study sample

              dim(Y)<-c(NStudySamples,length(Y))
             colnames(Y)<-colnames(X)

    }

##########################################
# Some computations before preparing the plot
##########################################

princomp(X, scale = FALSE)->X.PCA # computes the PC of the reference samples

X.PCA$loadings->M # gets the rotation matrix

(Y%*%M)[,1:2]->StudyCoordinates # computes the new coordinates of the study samples


if(NStudySamples==1){StudyCoordinates<-t(StudyCoordinates)} # forces StudyCoordinates to be a matrix

per<-eigen(cov(X))$values/sum(eigen(cov(X))$values) # this calculates the percentage of explained variance
 for each principal component

per<-round(per*100,2) # values are rounded to two decimal positions

# In this part colours in the plot must be chosen (if you have more than three populations)

mycolours<-c("orange","pink","skyblue2") # choose the colours you want to use for each population
 (considering that populations are in alphabetic order). A complete list of colour names can be obtained
 with the command colours() or with the help of the Chart of R colours available at http://research.stowers-
 institute.org/efg/R/Color/Chart/

colours<-as.character(factor(RefData$Population,labels=mycolours)) # population names are converted to the
 corresponding colour name


#####################
# The plot starts here
#####################

quartz() # opens a new graphic display window if you use MacOS
# windows() # is the alternative command for windows OS that opens a new graphic display window if you use a
 Windows PC (remove # here and replace # in front of previous line)

plot(X.PCA$scores[,1:2],col=colours,pch=20,main="put here your plot title",xlab=paste("PC1
 ",per[1],"%",sep=""),ylab=paste("PC2  ",per[2],"%",sep=""),cex=1.5)
# this plots the two first principal components. The plot title, pch and cex values can be changed

legend("topleft",legend=levels(factor(RefData$Population)),col=mycolours,pch=20,cex=0.5,y.intersp=1)
# this adds a legend to the plot. Its position can be changed using "topleft", "topright", "bottomleft" or
 "bottomright". Pch should match the one used in the plot.

### With the next set of commands it is possible to include the study individuals superimposed onto the
 previously plotted principal components graphic.

points(StudyCoordinates,col="black",pch=20,cex=2)
# this estimates the coordinates for the study individuals and plots them onto the previously generated
 principal component graphic. colour, pch and cex can be adjusted.

text(StudyCoordinates,as.character(StudyData[,1]),cex=1,pos=1)
# this adds study individual's descriptors to the plot
```

3.4.4 What Information Can Be Obtained from PCA?

PCA allows the exploration of data sets and shows proximity between individuals. In fact, it is possible to include a casework sample in the PCA plot generated for the reference populations helping to infer, through visual inspection, the most probable classification of that individual (Fig. 3).

3.5 Casework Example of a Custom Ancestry Inference: The 11-M Madrid Bomb Attack

In the 11-M Madrid bomb attack investigation, standard DNA analysis with STRs was supplemented with Y-filer and standard mtDNA analysis in most exhibits. But seven complete STR profiles, originating from five personal items together with a handprint on the handle of the bag containing an undetonated device, failed to match any of the suspects so these DNAs became the focus of specialist genotyping to analyze ancestry, specifically confined to the comparison of European with North African variability. This differentiation can be difficult to achieve for Y-chromosome and mtDNA due to differences in the scope and depth in the databases between European and North African data, so the 34-plex AIM-SNP set was chosen [24].

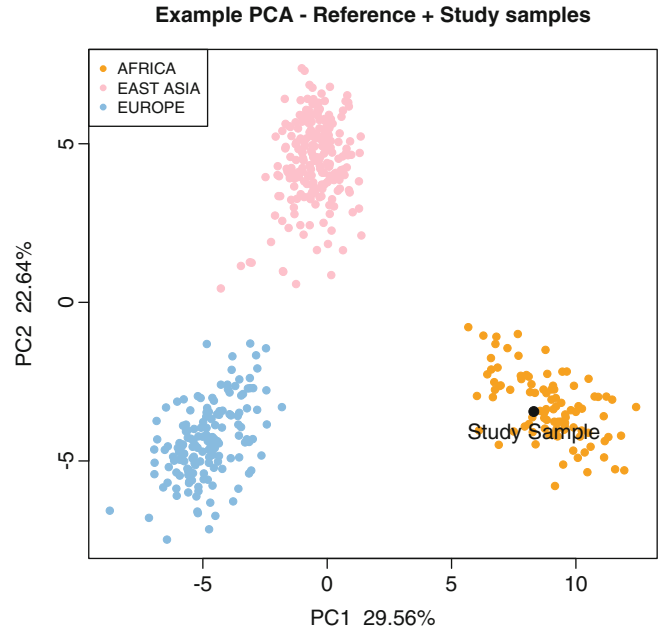


Fig. 3 PCA plot generated using the R script described in Subheading 3.4.3. Three population groups from the HGDP-CEPH panel of samples are used as reference data: Africa (*orange*), Europe (*blue*), and East Asia (*pink*). One study sample was plotted in the reference PCA (*black*). It is possible to infer that the study sample is likely to be African. Both reference and study samples genotypes are supplemented as text files in the online version of this chapter

The approach followed in this case is a good example of the integration in one analysis of the different techniques described in this chapter. Two training sets were made using 48 Moroccans and 48 Spanish from Madrid. Using *Snipper* a cross-validation assessed the accuracy (assignment success/error) and performance (range of likelihoods) of the training sets and to generate pairwise likelihood plots to assess patterns of possible admixture (Fig. 4). Such plots compare two ancestries and enable a simple comparison of the range of likelihoods observed in the unknowns alongside their closest parental population vs. another alternative population. The charts are made in Excel by converting *Snipper* likelihoods to whole numbers (using the =EXP formula in Excel), making each pairwise ratio (in this case, lk Moroccan/lk Spanish) and ranking values in descending order. Charting these with a logarithmic scale provides a simple visual check of the range of divergence between the populations compared as points with varying distances from the midline of 1 (equating to balanced odds of ancestry assignment to either population). The most distant points from the midline represent the strongest assignments. In populations without admixture, points are fully separated; when admixture occurs, a significant proportion of values are close to or cross the midline. Using *STRUCTURE*, admixture patterns were assessed in the training set. Some individuals, corresponding to likelihood ratios between 0.01 and 100, presented admixed ancestry. Considering this information, an area of uncertainty was defined, with individuals falling in that area not assigned to a particular population group.

When comparing PCA (*see* Fig. 2 in [24]), *STRUCTURE*, and *Snipper* results (Fig. 4), they were concordant for all case samples: three were classified as North African, one as European, and three were left unassigned. Those three unassigned profiles probably represent individuals with highly admixed parentage and genomic backgrounds: a reasonable scenario given the proximity of Southern Europe and North Africa.

One 34-plex assignment contradicted the uniparental analysis—although mtDNA and Y-chromosome markers routinely demonstrate informative geographic differentiation, this is not always true when recent gene flow has occurred or populations show strong sex bias (i.e., males are mainly from one population and females from another). The individual inferred to be European from uniparental data gave strong indications to be North African from the 34 SNP genotypes and was later identified by the investigation to be Algerian.

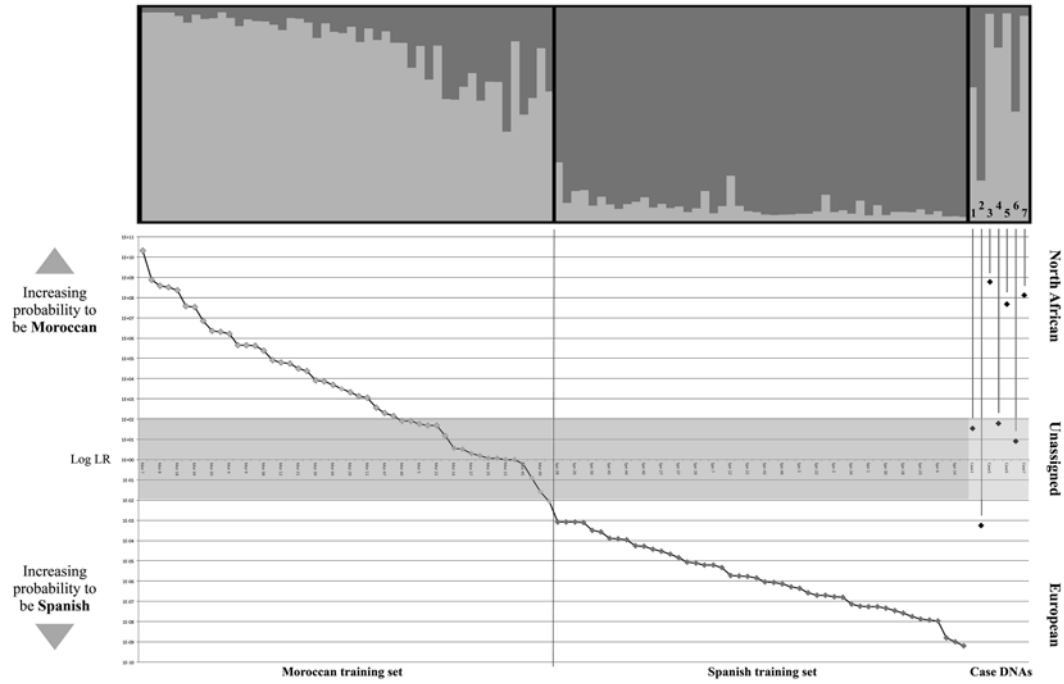


Fig. 4 11-M Madrid bomb attack *STRUCTURE* and *Snipper* analysis results. *STRUCTURE* analysis was performed to compare *Snipper* pairwise plots with an established alternative system of ancestry assessment. *STRUCTURE* runs were performed using the *admixture* ancestry model with 200,000 MCMC steps after a *burnin* of length 200,000. In the *Snipper* pairwise plot, samples are organized from most probable Moroccan to most probable Spanish, defining a separation from likelihood ratios represented on a logarithmic scale with values higher than 1 = higher probability North African and ratios smaller than 1 = higher probability European. Individuals in the *STRUCTURE* plot are in the same order as the *Snipper* pairwise plot, allowing direct comparison of both analyses. Some admixture patterns are present and this helped to establish an uncertainty area (ratios between 0.01 and 100) where individuals would have more probability of being misclassified, so the decision was to leave these unassigned

4 Notes

1. Of all databases included in *SPSmart*, 1000 Genomes and HapMap are of most utility as they have more markers and larger sample sizes (including admixed ancestry populations). This is particularly true for 1000 Genomes—the ENGINES browser allows scrutiny of SNP variation across the whole genome (down to a minor allele frequency of ~1 %) from Africans, Europeans, and East Asians previously used by HapMap. In contrast, the HGDP-CEPH (Human Genome Diversity Panel) samples 1050 individuals with wide currency

in population genetics studies [25, 26]. The geographic coverage is patchy in certain regions but all continents and all major genetic ancestry groups defined by studies of variability are represented. Smallest sampling is 28 Oceanians from two populations and just six San from Namibia. SPSmart provides freely downloadable genotype data from 650,000 SNPs (obtained with Illumina 650K arrays [27]) for each HGDP-CEPH sample in the Stanford University study of this panel.

2. Please note that, despite Excel software is referred as the one to be used, any spreadsheet software such as Numbers in MacOS or the free and open-source OpenOffice Calc or LibreOffice Calc can replace it.
3. Symmetrical base SNPs, comprising an A/C on one strand and a T/G on the other, are a particular problem and source of error when comparing genotypes generated on different platforms or listed in different databases. The *SPSmart* SNP*for*ID browser makes allowance for most base inversions by showing the HapMap (or other) allele frequency summary charts with different allele segments if these differ from the strand interrogated by the 34-plex and 52-plex extension primers. For example, rs2304925 is listed as a SNP*for*ID GT SNP but a HapMap AC SNP and this applies equally to GC or AT SNPs, e.g., rs10141763. The *SPSmart* help file provides a clear and carefully worded guide in the "Symmetrical bases" section. There are four symmetrical bases in the 34-plex: rs773658, rs10141763, rs1335873, and rs16891982. The last of these is the most informative SNP for differentiating component populations within Eurasia so it is particularly important to be clear about differences between SNaPshot and database allele calls for this marker.
4. Bayesian population analysis methods calculate a simple relationship between allele frequencies in a population and allele frequencies observed in the tested individuals. *STRUCTURE* analyzes differences in the distribution of genetic variants between populations through an iterative Bayesian algorithm that tries to group samples into clusters whose members share similar patterns of variation. Bayesian methods have the advantage of allowing the use of prior information about the samples to progress analysis. But the ability to differentiate populations in a sample set can be limited when a small number of samples and/or markers are used.
5. *STRUCTURE* uses an MCMC algorithm that starts by randomly assigning individuals to a predetermined number of K populations. Allele frequencies of each population are estimated

considering the individuals assigned to each population. Individuals are then re-assigned to populations taking into account the estimated frequencies for each population in a process repeated up to 10,000–100,000 times.

6. *Structure Harvester* implements the *delta K* method of Evanno et al. [23] to estimate the optimum *K* value [16, 17]. After running STRUCTURE, pass the zipped folder containing the results files (named *x_run_y_f*, where *x* represents the parameter set name and *y* the run number) to the *Structure Harvester* browser and click Harvest! to start the analysis. Conditions are: a minimum of three sequential *K* values should be included, with more than one replicate for each *K* value (same number of replicates for all *K* values). Posterior probability and *delta K* plots are available to download (Fig. 5)—the optimum *K* is usually the point with the highest *delta K* value or the one which immediately precedes the *Ln Pr*(*K*) plateau. This software is also useful as it automatically generates input files to run *CLUMPP*.
7. There are two basic ancestry models: *no admixture* and *admixture* [2]. The first is used if there is no prior knowledge about the origin of the populations under study or if there is a reason to consider each population as completely discrete. But because admixture between populations is a common characteristic, knowing the approximate median value of the ancestral population proportions for each individual and their populations of origin is very important for the characteriza-

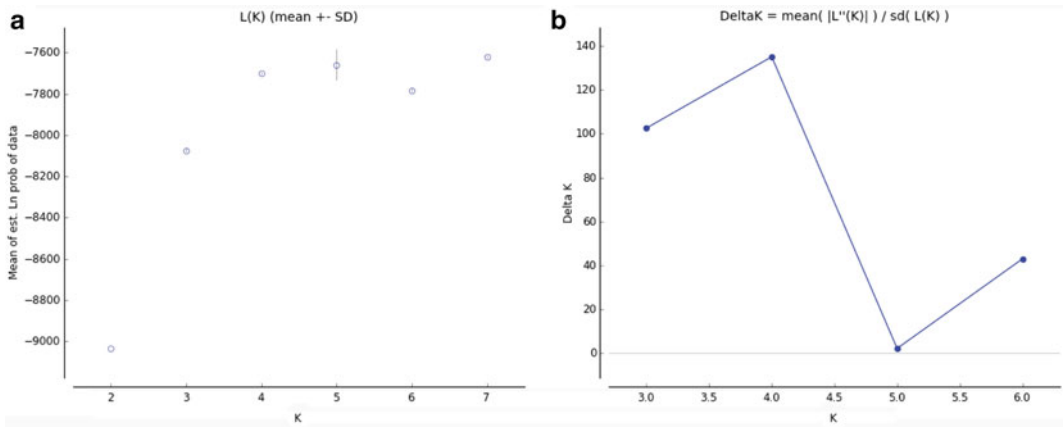


Fig. 5 Example of posterior probability and *delta K* plots obtained with *Structure Harvester* for the same analysis. In this case, the optimum *K* value is 4—the point where the *plateau* in the posterior probability starts with maximum *delta K* value

tion of a study population or, in a forensic context, a casework sample. In this case the *admixture* model is more appropriate. The *LOCPRIOR* option [19] can be used when there is additional sample characteristic data available, e.g., linguistic, geographical, cultural, or phenotypic information. The *LOCPRIOR* parameter is particularly informative when there are weak population structure signals—a situation that can result from using reduced number of markers, small sample sizes, or due to close relationships between populations. It has two main advantages: (1) generally it will not find population structure when this is not present; and (2) it can ignore location information when individual ancestry is not related with it. When admixture LD is present, the *linkage* model [18] (which is based on the *admixture* model) can be applied to obtain more accurate estimates of statistical uncertainty from use of linked markers. Population labels can be used to calculate the probability that each individual originates from the assumed population—individuals with low probabilities can be considered as migrants or having high co-ancestry. This option is included in the *USEPOPINFO* model [2] and should only be used when population labels are well defined beforehand and correspond almost exactly to the groups ultimately defined by the *STRUCTURE* results. The last model considers the specified information about the population of origin of a portion of individuals to help infer the ancestry of other samples with unknown origin: the *POPFLAG* model [2]. This option needs caution as selected samples are treated as the "reference" set (pre-assigned *POPFLAG=1*) meaning allele frequencies estimates are based on a reduced subset of samples and will directly affect the grouping of unknowns (pre-assigned *POPFLAG=0*). This model can be useful when grouping individuals/populations by comparison with very well-defined reference data—this option is particularly useful in the forensic context.

8. There are two allele frequencies models: *independent allele frequencies* and *correlated allele frequencies*. The first is used when frequencies are reasonably different in distinct populations—this implies that knowledge about the correlation level across the population is needed. The second assumes a non-independence level and offers more power to detect distinct populations that are closely related (e.g., Chinese and Japanese)—in the absence of high correlation levels, this model gives the same results as the *independent allele frequencies* model.

9. *STRUCTURE* does not assume a particular mutation process so the scale of the number of repeat units in STRs is not considered (only allele frequencies are important). For this reason there is no need to multiply all other alleles by 10 to compensate the transformation of intermediate alleles (19.3 = 193).
10. To confirm the number of markers and individuals select "Show data file format" showing total lines and columns. As an example, the data file format information for Table 1 would state: one line with m columns (m corresponds to the number of markers) and n lines with $m+6$ columns (four columns with prior information and two with extra information) with $n/2$ individuals (two lines per genotype).
11. When preparing the input file following the example presented in Table 1, there is no need to select the "special format" option because by default *STRUCTURE* assumes genotypes are arranged as two consecutive rows (diploid species) per individual. If both alleles are in the same line but in consecutive columns select "Data file stores data for individuals in a single line".
12. A *burnin* period of 10,000–100,000 is sufficient to observe convergence to an equilibrium point of parameters such as α—the relative admixture levels between populations. To check the variation of the parameter values go to the "Data plot" option in the simulation results window. When excessive variation is observed at the end of the *burnin* period, it is necessary to increase its length. To select an appropriate number of MCMC steps after the *burnin*, it is advisable to perform several simulations for each K value considering different lengths to see if the results are consistent—usually 10,000–100,000 MCMC steps are enough but to obtain precise posterior probability estimates longer simulations might be needed.
13. *STRUCTURE* performs individual analyses for each assumed population number from one up to a reasonable number for the sampling regime—at least three K values more than the number of expected population. If a *plateau* on the posterior probabilities is not reached, larger K values might be needed. Furthermore, clustering algorithms such as the one implemented in *STRUCTURE* can show stochastic variation from the simulations. To diminish their effect, several replicates for each K value should be made (at least three to five replicates advised).
14. Computational times can vary depending on the number of markers and samples to be analyzed, but also on the analysis parameters selected. As a point of reference, running a project

using the example input file supplemented with the online version of this chapter took approximately 3 h 45 min in a computer with a 2.7GHz Intel Core i7 processor. The project included two parameter sets: *admixture* and *admixture POPFLAG*—both were performed through 100,000 burnin steps, 100,000 MCMC repeats, three replicates from $K = 2$ to $K = 6$, and correlated allele frequencies.

15. Independently of the origin of differences between clustering results, a method to deal with replicate results is needed. *CLUMPP* uses replicates of the estimated membership proportion matrices for any *K* number—the result is a set of permutated matrices so that all the replicates have the best correspondence possible. It also generates a matrix that corresponds to the median of the permutated matrices. This is done for the population and individual proportions matrices. Currently the easiest way to prepare input files for *CLUMPP* is with *Structure Harvester* (*see* **Note 6**). Two files are needed: *.indfile* includes individual ancestry membership proportions tables from all replicates per *K* value and *.popfile* includes population ancestry membership proportions tables for all replicates per *K*. In both a blank line separates each table. If the input files are prepared manually take care with the sample name, which must be numeric; if alpha-numerical an error message appears. Both *.indfile* and *.popfile* files, together with *paramfile* and others, must be saved in the software folder together with the executable file. The *paramfile* includes important parameters that must be adjusted: *DATATYPE* defines which data is going to be considered for analysis (0 = individual; 1 = population); *K* is the number of clusters; *C* is the number of individuals or populations (depending on the selected *DATATYPE*); *R* is the number of replicates; *M* is the algorithm used. We recommend *M = 1* so all possible permutations are performed, but with large *K* or *R* values *M = 2* (10,000 random input repeats) is sufficient and for *K* values above 15 *M = 3* is advisable. Metric *S* is the pairwise matrix similarity statistic and we recommend the standardized *G′* (select *S = 2*). It is important to note that to obtain a population and individual mean matrices, two runs are required, adjusting the *.output* file name (no name change overwrites the first run), *DATATYPE* and *C* in between runs. In the Windows version, just execute the *CLUMPP.exe* file—a *cmd* window opens showing the progress of the simulation. In the MacOS version execute the software through the terminal command line: change the directory to the *CLUMPP* folder location (type *cd* > drag the folder into the terminal > ENTER) and then type *./CLUMPP* > ENTER to run the software.

16. A convenient way of visualizing *STRUCTURE* results (especially for $K > 3$) is to show each individual as a straight segment divided into K colors that represents the estimated ancestry membership proportions. *STRUCTURE* gives such bar plots but their format cannot be changed and they only present replicate results for one K value. *Distruct* offers a great variety of options to generate more informative cluster plots. As with *CLUMPP*, *distruct* uses a set of files stored in the same folder of the executable file. The input files include the population Q-matrix (*.popq* file) and the individual Q-matrix (*.indivq* file) obtained directly from *STRUCTURE* (in the case of a single K replicate) or from *CLUMPP*. Files: *.names* and *.languages* define the labels above and below the plot. Both files have the same format: in each line write the population numeric code, space, and preferred name. To define the color of each cluster open the *.perm* file—with a minimum K lines each defining a color. Colors are assigned to each cluster and not to each predefined population, i.e., if population 1 appears in orange and this population is defined in cluster 3 in *.perm*, define "3 orange" and not "1 orange". The *drawparams* file has several modifiable parameters, notably: K, *NUMPOPS* (number of populations) and *NUMINDS* (number of individuals). Remaining parameters adjust graphical aspects of the plot: letter size, distance between text and plot, height of the plot, thickness of the columns representing individuals, thickness of the contour lines, horizontal/vertical orientation, and others (Fig. 6). The "//" symbol indicates that the following text is a comment describing the parameter function and that it will not be used by the software. When computations are complete a *PostScript* (*.ps*) file containing the plot is created To visualize the plot with Windows, specific software such as *GhostView* (freely available online), Acrobat Distiller or Illustrator is needed. With MacOS plots can be opened with preview and exported as a pdf.
17. A new option has been added to Snipper: the ability to apply or not apply the HWE principle. The assumption of HWE when not valid may result in inaccurate genotype frequency estimates and, in turn, an inaccurate classification.
18. Each missing genotype is entered as two Ns per SNP, so only ACGTN characters are permitted. Blank spaces are ignored. Format errors from incorrect bases (either due to incorrect SNP order or inverted bases) are flagged by Snipper with a warning for the relevant SNP position(s).
19. In the supplementary Excel input file included in the online version of this chapter, an example concatenated profile is given. This individual will be classified as African—use Option

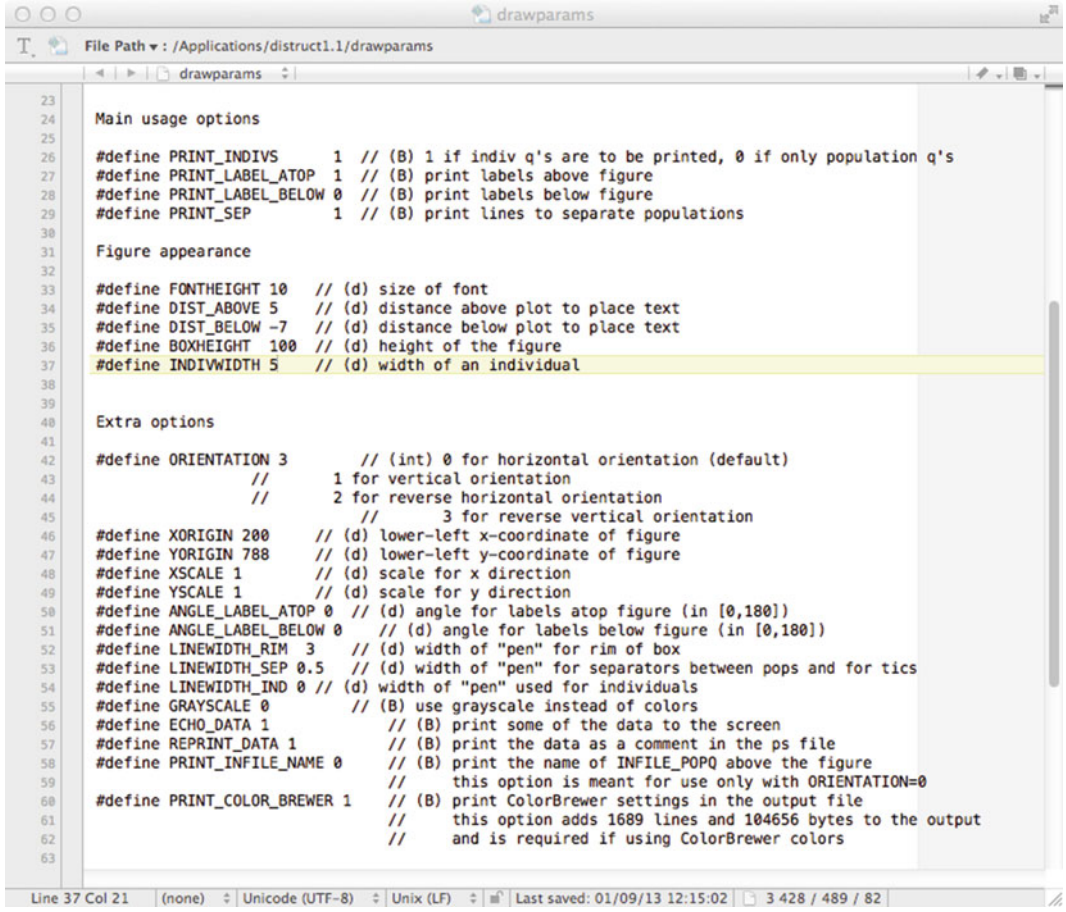


Fig. 6 Example of *distruct* parameters. Considering an output file resembling an A4 sheet lying horizontally (longer side down), and depending on the number of samples and the desired effect, it is worth taking advantage of available space. Reverse horizontal orientation uses the longest side of the virtual sheet. Changing the *XORIGIN* and *YORIGIN* values also helps—for example, *XORIGIN* = 200 moves the plot away from the margin of the sheet and *YORIGIN* = 788 leaves just enough space to separate the plot from the margin without leaving much unused space. The individual bar width (*INDIWIDTH*) can then be adjusted to an appropriate value that allows all the individuals to be included in the virtual sheet

1 or Option 2 in *Snipper* as described above (for Option 2 use the supplementary Excel file as population data input—remove the last two samples (unknown ancestry) and the last column; adjust the number of individuals in cell 1A to 479 and the number of populations in cell 1C to 3). Note that -log(LIKELIHOOD) values are returned, so lower values are better.

20. Divergence is calculated based on the number of populations included in the comparison. For example, on the fixed training

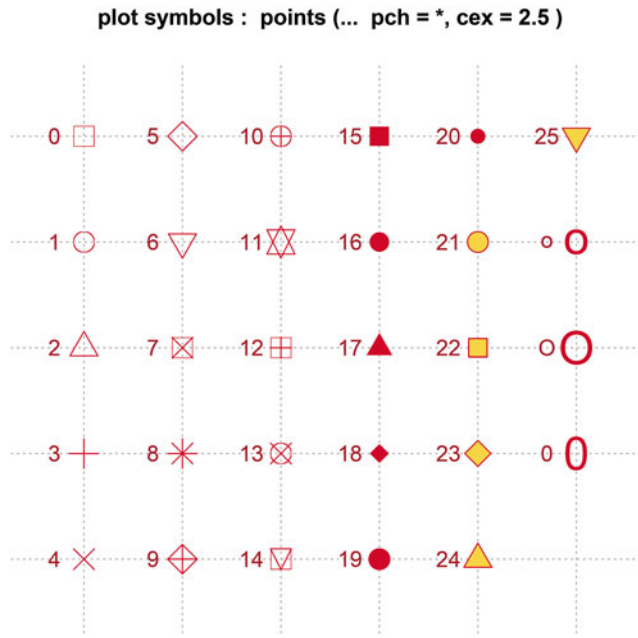


Fig. 7 Symbols available in *R* to define the points shown in plots (*pch* command)

set option, divergence will be calculated based on 3, 4, or 5 groups depending on the option selected in Step 2.

21. When generating PCA plots, it is possible to change graphical parameters to user needs. In the *R* script, command lines are included so the "Population" information can be used to define color of the symbols in plots. In this case, population names are transformed into color names, in population alphabetical order and not input order (a complete list of color available in *R* can be obtained with the command *colours()* (alternatively *colors()*) or with the help of the *Chart of R colours* available at http://research.stowers-institute.org/efg/R/Color/Chart/). The symbol shape can also be changed—for information on available symbols use the *pch* help page by typing *?pch* in R console (Fig. 7). The *cex* command changes the size of the points.
22. If you are using *SNPassoc* package for the first time, you need to install it in *R*. Two options are available: (1) install it from the Package installer option in the *R* console; or (2) download the package zip file from http://cran.r-project.org/web/packages/SNPassoc/index.html and perform a local zip file installation.

References

1. Jobling M, Hollox E, Hurles M et al (2014) Human evolutionary genetics: origins, peoples & disease, 2nd edn. Garland Science - Taylor & Francis Group, New York
2. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2): 945–959
3. Phillips C, Salas A, Sánchez JJ et al (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. Forensic Sci Int Genet 1(3–4): 273–280
4. Fondevila M, Phillips C, Santos C et al (2013) Revision of the SNPforID 34-plex forensic ancestry test: assay enhancements, standard reference sample genotypes and extended population studies. Forensic Sci Int Genet 7(1):63–74
5. Pereira R, Phillips C, Pinto N et al (2012) Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. PLoS One 7(1):e29684
6. Jolliffe I (2002) Principal component analysis. Springer, New York
7. R Development Core Team (2011) R: a language and environment for statistical computing. http://www.r-project.org
8. Phillips C (2009) SNP databases. In: Komar AA (ed) Single nucleotide polymorphisms, vol 578, Methods in molecular biology. Humana, New York, pp 43–71
9. Amigo J, Salas A, Phillips C et al (2008) SPSmart: adapting population based SNP genotype databases for fast and comprehensive web access. BMC Bioinformatics 9:428
10. Amigo J, Phillips C, Lareu MV et al (2008) The SNPforID browser: an online tool for query and display of frequency data from the SNPforID project. Int J Legal Med 122(5): 435–440
11. Amigo J, Salas A, Phillips C (2011) ENGINES: exploring single nucleotide variation in entire human genomes. BMC Bioinformatics 12:105
12. Amigo J, Phillips C, Salas A et al (2009) pop.STR—an online population frequency browser for established and new forensic STRs. Forensic Sci Int Genet Suppl Ser 2(1):361–362
13. Rosenberg NA, Pritchard JK, Weber JL et al (2002) Genetic structure of human populations. Science 298(5602):2381–2385
14. Rosenberg NA, Li LM, Ward R et al (2003) Informativeness of genetic markers for inference of ancestry. Am J Hum Genet 73(6): 1402–1422
15. Kalinowski ST (2011) The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity 106(4):625–632
16. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14(8):2611–2620
17. Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4(2):359–361
18. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164(4):1567–1587
19. Hubisz MJ, Falush D, Stephens M et al (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9(5):1322–1332
20. Porras-Hurtado L, Ruiz Y, Santos C et al (2013) An overview of STRUCTURE: applications, parameter settings and supporting software. Front Genet 4:98
21. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14):1801–1806
22. Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4(1):137–138
23. Gonzalez JR, Armengol L, Sole X et al (2007) SNPassoc: an R package to perform whole genome association studies. Bioinformatics 23(5):644–645
24. Phillips C, Prieto L, Fondevila M et al (2009) Ancestry analysis in the 11-M Madrid bomb attack investigation. PLoS One 4(8):e6583
25. Cann HM, de Toma C, Cazes L et al (2002) A human genome diversity cell line panel. Science 296(5566):261–262
26. Rosenberg NA (2006) Standardized subsets of the HGDP-CEPH human genome diversity cell line panel, accounting for atypical and duplicated samples and pairs of close relatives. Ann Hum Genet 70:841–847
27. Li JZ, Absher DM, Tang H et al (2008) Worldwide human relationships inferred from genome-wide patterns of variation. Science 319(5866):1100–1104

Chapter 20

Species Determination: The Role and Use of the Cytochrome *b* Gene

Adrian Linacre and James Chun-I Lee

Abstract

The illegal trade in ivory and rhino horn has led to a catastrophic decline in elephant and rhino populations worldwide. These iconic high profile species are one part of the illegal trade in wildlife products that threatens these species and the ecological balance in the habitats in which they live. Identification of the species present is required to determine that the trade is illegal and contrary to national legislation. This chapter details a robust DNA technique using part of the cytochrome *b* gene on the mitochondrial genome that will work on poor quality samples such as powdered horn or ivory products including statues and carvings. An appropriate DNA extraction technique is required to obtain at least 1 ng of DNA from which the amplification of part of the cytochrome *b* gene using universal primers is performed. This produces a fragment of 486 bp in size which can be sequenced using standard technologies. The resulting sequences are then aligned to voucher specimens or sequences on reliable databases. Analyses of the data should lead to confident species identification.

Key words CITES, DNA, Mitochondria, Cytochrome *b*, Species identification, Wildlife

1 Introduction

The trade in endangered species is an acute problem in many parts of the world affecting many rare animal and plant species. High profile examples include the body parts of tiger and rhino, which play a part in spurious medicines, and the trade in elephant ivory. The value of the illegal trade in wildlife is illustrated by rhino horn, which is now estimated to be worth between $65,000 and $100,000 USD per kilogram making it more expensive than gold [1]. Ivory is equally sought after leading to the slaughter of elephants and tigers are considered a living pharmacy contributing to a dramatic decline and extinction in many of their previous habitats. Additionally, the illegal trade in endangered exotic birds, reptiles, and hard woods leads to wildlife crime being highly lucrative, but disastrous for the species traded (e.g., *see* ref. 2). The Convention on International Trade in Endangered Species of Wild Flora and

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0_20,

Fauna (CITES) lists a growing and depressing number of species on the verge of extinction in the wild (see www.cites.org). As so many of these animal species are traded illegally, often for their skins or for supposed medicinal qualities, it is the role of laboratories to determine if seized samples are from a protected species.

This chapter focuses on DNA sequence techniques as they work effectively on trace material. The choice of DNA locus is key to the success of the test and there are certain criteria that must be met. It is necessary to use a locus that is sufficiently conserved such that all members of the same species have a very similar DNA sequence; the locus must be sufficiently polymorphic such that members of different, but closely related species, can be distinguished; the locus must be on a fragment of DNA that can be amplified from poor quality samples; and it must be possible to analyze the locus from a wide range of species using the same protocol. Two loci on the mitochondrial genome have been used routinely; these are the cytochrome *b* (cyt *b*) [3–12] and cytochrome oxidase 1 (COI) [13–16] genes. Other loci can be used and while this chapter focuses on the cyt *b* locus, the same methodology can be applied to any gene locus of comparable size.

The cyt *b* gene is part of a complex of genes involved with oxidative phosphorylation [17]. The complete cyt *b* gene is approximately 1140 bp and encodes a protein of 380 amino acids in length. It may not always be possible to amplify such a large fragment from processed samples. A range of primers have been published that amplifies a 486 bp fragment of this locus from any mammalian DNA (Table 1). If performing species identification using another locus it is essential that the target sequence has sufficient interspecies variation and that there are comprehensives DNA sequence databases available for comparison purposes (*see* **Note 1**). The primers used in this chapter are designed to amplify a 486 bp product of which 402 bp are used for sequence comparison. The structure of the cyt *b* gene is shown in Fig. 1.

Table 1
Primers used in amplifying part of the Cyt *b* gene

Location	Sequence (5′–3′)	Size (bp)	Reference
L14841 H15149	AAAAAGCTTCCATCCAACATCTCAGCATGATGAAA AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA	308	[19]
L14724 H15149	CGAAGCTTGATATGAAAAACCATCGTTG AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA	425	[17]
L14816 H15173	CCATCCAACATCTCAGCATGATGAAA CCCCTCAGAATGATATTTGTCCTCA	358	[20]
L14724 H15149	CGAAGCTTGATATGAAAAACCATCGTTG AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA	425	[5]

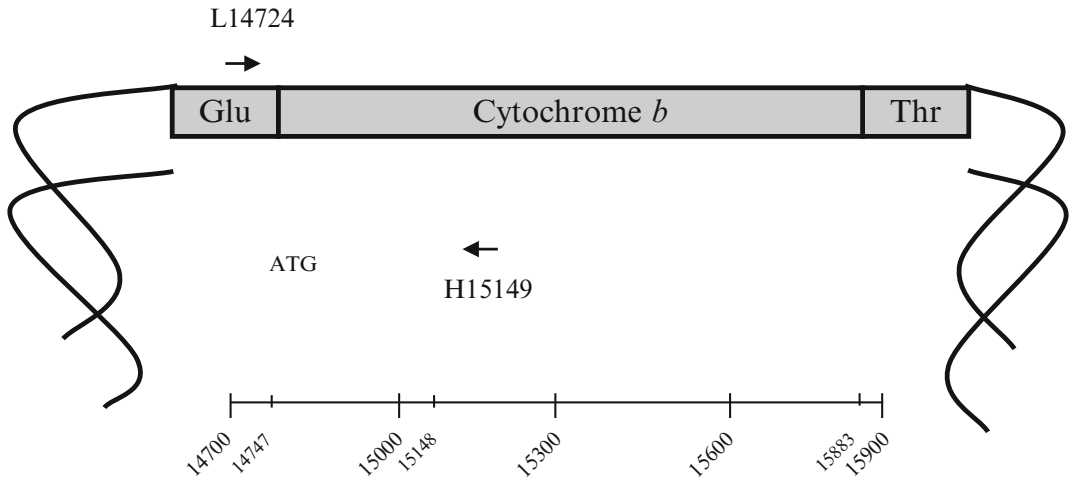


Fig. 1 The position of the primers 14724 and H15149 on mitochondrial DNA is shown. Numbering is according to the human mtDNA sequence [18]

A list of the primers used in amplifying a section of the cyt *b* gene for the purpose of species identification is shown in Table 1. These primer sets are listed to show those that have been published for the purpose of species identification. The primer sets listed on the last two rows of the table are those validated for a range of mammalian species specifically for the purpose of species identification. Note that the position of the primers at the 3′ end is based on the human DNA sequence [18] and will be different for other mammalian species.

2 Materials

1. L14724 (7.5 μM): 5′-CGAAGCTTGATATGAAAAACCATCGTTG-3′.
2. H15149 (7.5 μM): 5′-AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA-3′.
3. AmpliTaq Gold DNA Polymerase (Applied Biosystems, CA, USA) (5 U/μL).
4. AmpliTaq Gold PCR Buffer (Applied Biosystems, CA, USA) (10×).
5. dNTPs (10 mM each of dATP, dCTP, dGTP, and dTTP diluted in sterile H_2O).
6. DNA Terminator Sequencing Kit (ABI PRISM™ BigDye™ Cycle Sequencing Kit, Applied Biosystems, CA, USA).
7. Quick Spin Columns (Boehringer-Mannheim, Mannheim, Germany).
8. Template Suppression Reagent (Applied Biosystems, CA, USA).

9. Thermal cycler.
10. PRISM 310 Genetic Analyzer (Applied Biosystems, CA, USA).
11. Vacuum Evacuator.
12. MicroAmp Reaction Tubes.
13. EXO-Sap (Illustar™, GE Healthcare, Victoria, Australia).

3 Methods

3.1 PCR Amplification of Part of the Cyt *b* Locus

1. Aliquot approximately 1 ng of DNA into a sterile 0.2 mL tube (*see* **Notes 2** and **3**).
2. Add 1 μL of each primer (final 0.15 μM of each) (*see* **Note 4**).
3. Add 5 μL of PCR Buffer.
4. Add 5 μL dNTPs.
5. Add 0.5 μL of AmpliTaq Gold.
6. Add water to 50 μL.
7. Amplify using the following conditions: 95 °C for 11 min, followed by 35 cycles of 95 °C for 45 s, 50 °C for 45 s, and 72 °C for 90 s, followed by 72 °C for 30 min (*see* **Notes 5** and **6**).

3.2 Confirmation of PCR

Samples can be separated on an agarose gel to confirm the success of the amplification. A 2 % agarose gel should be cast in 1× TAE with 1 μg/mL ethidium bromide. A small size gel is normally sufficient requiring 50 mL of agarose mix.

1. Add 1 g agarose to 50 mL TAE.
2. Heat the agarose in a microwave to dissolve the agarose completely and allow to cool to 55 °C before casting the gel.
3. Once the gel has solidified, run 5 μL of each sample along with a size marker for 20 min at around 80 V before visualizing the bands on a UV transilluminator (245 nm). The PCR products should be approximately 486 bp in length. An example of a successful amplification is shown in Fig. 2. *See* **Note 7** if no PCR product is present.

3.3 Sequencing of PCR Products

3.3.1 PCR Cleanup

The PCR products can be directly sequenced. As the size is 486 bp the complete sequence can be determined in one reaction, providing the PCR product is sufficiently pure. The primers used in the PCR can be used in the sequencing reaction, but the buffer is different for sequencing compared to PCR and the unused dNTPs need to be removed. A standard means is to use an Exo-Nuclease to digest the single-stranded primers, and Shrimp Alkaline Phosphatase (SAP) to digest the free dNTPs. The protocol for this procedure is called commonly Exo-SAP as the two enzymes can be used in combination. The method below is for one reaction but if

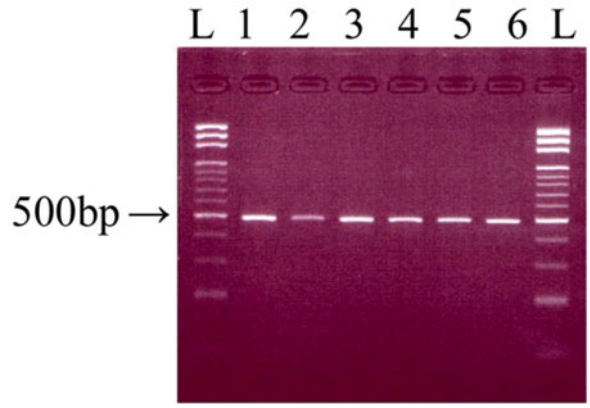


Fig. 2 An agarose image of PCR products after amplification of part of the cyt *b* gene. *Lane L* is the 100 bp ladder; *lanes 1–6* are PCR products from human, macaque, Asian elephant, mouse, white rhino, and pig respectively. The expected size of the product is 486 bp

sequencing with both the forward and reverse primer separately (i.e., sequencing both strands) then simply double the volume of this Exo-SAP reaction.

1. Incubate 5 μL of PCR product with 2 μL of Exo-SAP (typically 1 U of enzyme) at 37 °C for 15 min.
2. Transfer the tubes to a heater set at 80 °C and incubate for 15 min to denature both enzymes. The solution is now ready for the addition of a single primer used in a sequencing reaction.

3.3.2 Cycle Sequencing Method

The method of DNA sequencing is a standard dye labeled terminator method. A commercial kit is used (ABI PRISM™ BigDye™ Terminator) as this method has been found to be successful. The optimum amount of template DNA is approximately 50 ng, although DNA concentrations of between 10 and 80 ng will still produce good quality sequence data. The primers used in the DNA sequence reaction are the same as those used in the amplification.

The DNA sequence of both strands should be determined and compared to ensure that there are no differences due to the sequencing reaction.

1. To two separate tubes, aliquot 7 μL of template solution containing the purified PCR product (*see* **Note 8**). Make up to 10 μL with sterile H_2O.
2. Add 8 μL sequencing reaction mix to the tube.
3. Add 2 μL of primer L14724 to one tube and primer H15149 to the other. The final concentration of each primer should be 0.25 μM.

4. Seal the tubes and place in a thermal cycler.
5. Perform cycle sequencing reactions using the following conditions: 96 °C for 1 min, followed by 25 cycles of 96 °C for 10 s, 50 °C for 5 s, and 60 °C for 4 min, followed by a rapid thermal ramp to 4 °C and hold until ready to purify.

3.3.3 Separation of Sequenced Fragments

1. The extract from the DNA sequencing reaction must be treated to remove the unincorporated primers. A number of methods are available to do this, and most use a simple spin column. Pipette the fluid containing the sequenced DNA into the top basket and spin for 10 min at 13,000 × *g* (check the manufacturer's protocol as this might vary depending on the tube used) to remove the unincorporated primers and free bases. The final volume will be ideally around 5 μL.
2. Pipette this volume into a new 1.5 mL tube.
3. Place the samples in a vacuum evacuator and allow the samples to dry (*see* **Note 9**).
4. Resuspend the dried pellet in 25 μL of Template Suppression Reagent.
5. The samples are now ready for separation. This is preformed almost exclusively by capillary electrophoresis using any suitable platform, for example the ABI 3500.

3.4 Analysis of Results

1. The expected size of the sequence products should be 486 bp based on the length in bp of the cyt b gene (*see* **Note 10**).
2. Remove sequence from the 5′ end of the L14724 primer to mitochondrial position 14746 so that the sequence starts at the start of the cyt *b* gene (position 14747) (*see* **Note 11**). This should result in sequence lengths of 402 bp if the original product was 486 bp.
3. Align the sequence data from each of the two primers used to determine a consensus DNA sequence for the gene sequence. Alignments can be performed using a range of software (such as Sequencher and Geneious). DNA sequencing should ideally be performed in both directions. Any anomalies or bases that are not identified clearly on one strand should be confirmed by sequencing of the complementary strand.
4. Export the DNA sequences to a software program such as GCG or MEGA (*see* **Note 12**).

3.5 Analyses of DNA Data

1. If the full 402 bp sequence is determined and if all the bases are designated with confidence, then the sequence should be compared to known DNA sequences on DNA sequence databases or from a voucher specimen (Fig. 3).
2. The percentage similarity between the unknown and known sequences should be recorded. The example shown in this

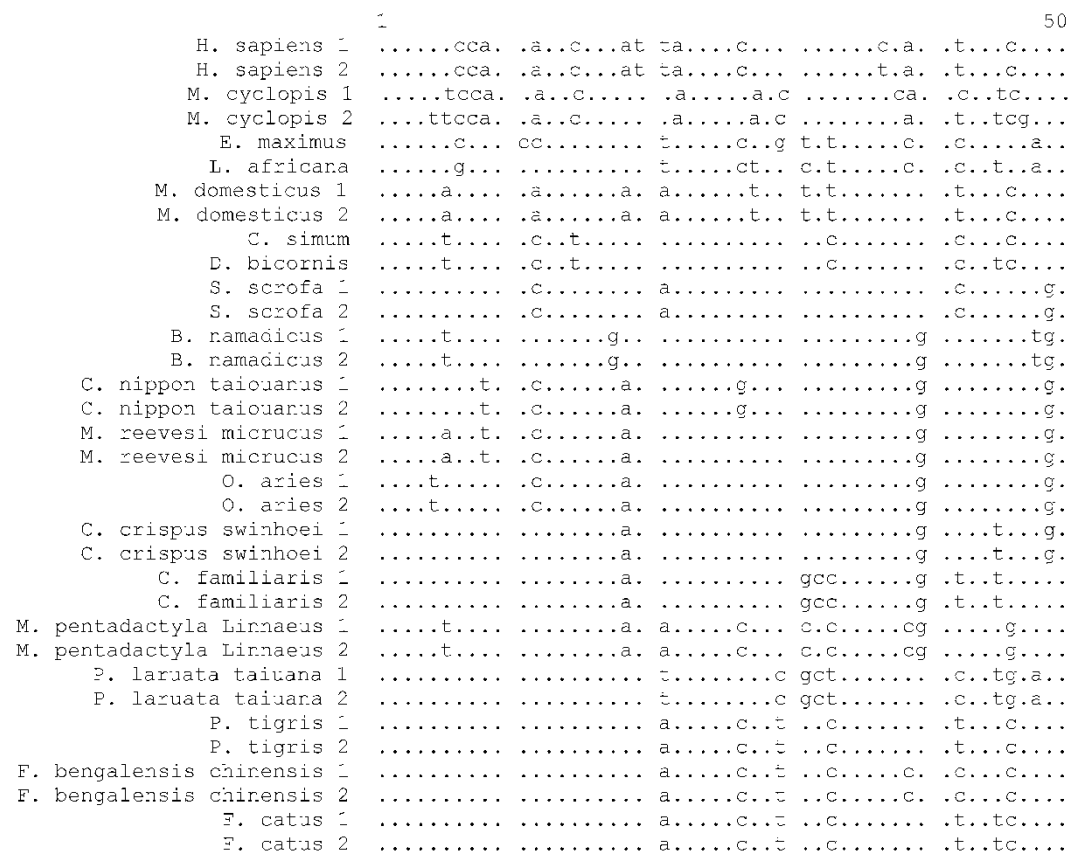


Fig. 3 Alignment of the first 50 bp of the PCR products from the cyt *b* gene from 19 different mammalian species. The sequence alignment was by the PileUp program of the GCG computer package and the consensus sequence was determined using the Pretty program. "." indicates the same base as the consensus sequence

chapter analyses the DNA sequence of a mammalian species where it is known that the minimum interspecies variation is 2.5 % and maximum intraspecies variation is 1.5 % [21].

3. Based on the above assumption, there should be no more than six bases different over the 402 bp aligned between the known and unknown DNA sequences. If such a similarity is recorded, there is a 99.9 % probability that the unknown is from the same species as the known (*see* **Note 13**).

4 Notes

1. Both cyt *b* and COI have extensive coverage in the DNA databases such as GenBank (www.ncbi.nlm.nih.gov/genbank) and for COI there is also the Barcode of Life Consortium (www.barcodinglife.com).

2. DNA should be extracted, purified, and quantified using a standard method. The use of chelating agents such as Chelex® is simple and safe but will not always produce DNA free from inhibitors. Solid phase extraction using a silica membrane (e.g., QIAamp micro kit from Qiagen) or a resin (e.g., DNA IQ from the Promega corporation) have proved a successful means of extracting DNA from a wide range of tissue types. DNA can also be extracted from bone using the salt chloroform method. An extraction negative should be performed at the same time as the unknown sample. Analysis of any reference sample should be performed either in a different laboratory or at a different time; this minimizes the opportunity for any contamination. An out-group control DNA can be used at the same time as unknown DNA. A good control DNA is from a domestic or farmed animal unlikely to be the same species as the unknown. Cow (*Bos namadicus*) or Mouse (*Mus domesticus*) DNA is ideal and normally available.
3. This protocol uses 1 ng of total DNA extracted. As the DNA to be amplified is on the mitochondrial genome, there are many more copies of this DNA relative to nuclear counterparts and therefore significantly less (100 pg) template DNA should produce a PCR product with good quality DNA. It may be however that the DNA is highly degraded and hence adding more than the suggested 1 ng is needed.
4. The primers used in this method amplify a 486 bp fragment. This primer set has been found to be useful in the amplification of products from highly degraded samples and is also validated for species identification. The full cyt *b* gene locus can be amplified from good quality samples using the primer sets described in ref. 5.
5. If more than three samples are being prepared it is better to create a master mix. For each sample add 1.1 μL of each primer, 5.5 μL buffer, 5.5 μL of dNTPs, 0.5 μL of enzyme, and make with a marginal excess of dH_2O to a volume of 50 μL per sample. Aliquot the appropriate volume into separate tubes then add 1–10 ng of template DNA.
6. The cycling parameters have been specified using the Applied Biosystems 9600. It is possible that the parameters will need to be altered if another thermal cycler is used.
7. If no PCR product can be seen on the gel then it is likely that there will be insufficient DNA to act as a template of the DNA sequencing. Initially check to see if the positive control has worked. If performing a master mix (*see* **Note 5**) then a product for the positive control indicates that the reaction should have worked if there was good quality DNA and at sufficient

quantity. It is most likely that more DNA is needed. Increasing the number of cycles past 35 is not recommended.

8. The quantity of DNA template required to generate optimal sequence data is 50 ng. This should be diluted into a volume of 10 μL. Adequate sequence data can be obtained from as low as 20 ng and as high as 100 ng of template DNA. It is advised to dilute, or concentrate, the sample to between 20 and 100 ng in a volume of 10 μL.
9. A vacuum pump is the normal method for drying the sequence products, but if unavailable it is possible to place the open reaction tubes used to collect the purified sequence products in a heating block with water filling the holes of the heating block. Leave the heating block at 95 °C to remove the fluid.
10. Most mammalian species will produce a fragment of 486 bp using the primers listed in this chapter. There are exceptions including the African lion (*Panthera leo*) where there are 21 bases that are deleted resulting in a fragment size of 464 bp.
11. The human consensus sequence starts at position 14747 with the sequence 5′-ATGACCCCAAT-3′. Edit the sequence to the 5′ side of this particular DNA sequence.
12. A range of free and commercial software programs can perform sequence alignments. Sequencher and Geneious are commercially available. MEGA performs a range of sequence alignments and is of more use once a verified sequence is confirmed.
13. A match between an unknown sample and a reliable sequence from a known species with a 100 % similarity over 400 bp of the cyt *b* locus indicates that either the unknown is from the same species as the known; the unknown is from another species on the database but there is an exceptionally large amount of interspecies variation (this is very unlikely given the data in ref. 21); or the unknown is from a species not on the database (this depends on the coverage of the database).

References

1. Linacre AM, Tobe SS (2013) Wildlife DNA analysis. Wiley, Chichester, UK
2. Wilson-Wilde L, Norman J, Robertson J, Sarre S, Georges A (2010) Current issues in species identification for forensic science and the validity of using the cytochrome oxidase I (COI) gene. Forensic Sci Med Pathol 6:233–241
3. Coghlan ML, White NE, Parkinson L, Haile J, Spencer PBS, Bunce M (2012) Egg forensics: an appraisal of DNA sequencing to assist in species identification of illegally smuggled eggs. Forensic Sci Int Genet 6:268–273
4. Hsieh HM, Huang LH, Tsai LC, Liu CL, Kuo YC, Hsiao CT, Linacre A, Lee JCI (2006) Species identification of Kachuga tecta using the cytochrome b gene. J Forensic Sci 51:52–56
5. Hsieh H-M, Chiang H-L, Tsai L-C, Lai S-Y, Huang N-E, Linacre A, Lee JC-I (2001) Cytochrome b gene for species identification of the conservation animals. Forensic Sci Int 122:7–18

6. Hsieh H-M, Huang L-H, Tsai L-C, Kuo Y-C, Meng H-H, Linacre A, Lee JC-I (2003) Species identification of rhinoceros horns using the cytochrome b gene. Forensic Sci Int 136:1–11
7. Lee J, Hsieh HM, Huang LH, Kuo YC, Wu JH, Chin SC, Lee AH, Linacre A, Tsai LC (2009) Ivory identification by DNA profiling of cytochrome b gene. Int J Leg Med 123: 117–121
8. Lee JCI, Tsai LC, Liao SP, Linacre A, Hsieh HM (2009) Species identification using the cytochrome b gene of commercial turtle shells. Forensic Sci Int Genet 3:67–73
9. Lee JCI, Tsai LC, Yang CY, Liu CL, Huang LH, Linacre A, Hsieh HM (2006) DNA profiling of shahtoosh. Electrophoresis 27:3359–3362
10. Lee JC-I, Tsai L-C, Huang M-T, Jhuang J-A, Yao C-T, Chin S-C, Wang L-C, Linacre A, Hsieh H-M (2008) A novel strategy for avian species identification by cytochrome b gene. Electrophoresis 29:2413–2418
11. Linacre A (2006) Application of mitochondrial DNA technologies in wildlife investigations—species identification. Forensic Sci Rev 18:1–8
12. Linacre A, Lee JC-I (2005) Species identification: the role of the cytochrome b gene. In: Carracedo A (ed) Forensic DNA typing protocols. Humana, Totowa, NJ, pp 45–53
13. Dawnay N, Ogden R, McEwing R, Carvalho GR, Thorpe RS (2007) Validation of the barcoding gene COI for use in forensic genetic species identification. Forensic Sci Int 173:1–6
14. Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc B Biol Sci 270:313–321
15. Hebert PDN, Ratnasingham S, de Waard JR (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Lond B Biol Sci 270:S96–S99
16. Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. PLoS Biol 2:1657–1663
17. Irwin D, Kocher T, Wilson A (1991) Evolution of the cytochrome b gene of mammals. J Mol Evol 32:128–144
18. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147
19. Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial-DNA evolution in animals—amplification and sequencing with conserved primers. Proc Natl Acad Sci U S A 86:6196–6200
20. Parson W, Pegoraro K, Niederstatter H, Foger M, Steinlechner M (2000) Species identification by means of the cytochrome b gene. Int J Leg Med 114:23–28
21. Tobe SS, Kitchener AC, Linacre AMT (2010) Reconstructing mammalian phylogenies: a detailed comparison of the cytochrome b and cytochrome oxidase subunit I mitochondrial genes. PLoS One 5:e14156

INDEX

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420,
DOI 10.1007/978-1-4939-3597-0, © Springer Science+Business Media New York 2016

L

M

N

O

P

Q

R

S

T

V

W

Y