Københavns Universitet

A 48-plex autosomal SNP GenPlex™ assay for human individualization and relationship testing

Tomas Mas, Carmen; Børsting, Claus; Morling, Niels

Springer Protocols

Antonio Alonso   *Editor*

DNA
Electrophoresis
Protocols for
Forensic Genetics

Humana Press

METHODS IN MOLECULAR BIOLOGY™

DNA Electrophoresis Protocols for Forensic Genetics

Edited by

Antonio Alonso

Instituto Nacional de Toxicología y Ciencias Forenses, Servicio de Biología, Madrid, Spain

☼ Humana Press

Editor
Antonio Alonso, Ph.D.
Instituto Nacional de Toxicología y Ciencias Forenses
Servicio de Biología, Madrid, Spain
a.alonso@mju.es

Preface

At present, most of the validated forensic DNA profiling procedures are based on high-resolution and high-throughput capillary electrophoresis separation and detection systems of PCR amplicons obtained from DNA genomic markers with different inheritance patterns (autosomal, Y-chromosome or X-chromosome linked, and mitochondrial DNA). A high degree of standardization has been achieved in the field of typing procedures for PCR products obtained with fluorogenic labeled primers, separated by capillary electrophoresis, and detected by the fluorescence emission induced by laser on a CCD camera. Current systems have acquired a high degree of accuracy and precision in the identification of allelic variants (with a resolution of one base), and also an improvement in the sensitivity of PCR amplicon detection (with reproducible signals even from picograms of DNA template). Another advantage of current capillary electrophoresis systems (with 5–6 multichannel detection technology) is a greater capacity for simultaneous detection of multiple markers. It can be detected up to 15–26 short tandem repeats (STRs) multiallelic markers, perform simultaneously about 50 biallelic single nucleotide polymorphisms (SNPs) or detect 4-dye-terminator sequencing reactions in a single capillary electrophoresis injection.

Human forensic identification is currently performed using commercial kits for STR multiplexing. Basically there are two sets of standardized STR markers in the criminal DNA databases around the world: the European standard set of 12 STR markers and the USA CODIS standard of 13 STR markers. They form together a standard of 18 STR markers in total. Development of mini-STR markers and their recent incorporation into commercial kits have improved the application of these markers in severely degraded DNA samples.

Other source of human genetic variation used in forensic casework is the analysis of sex chromosome markers. The Y-chromosome STR haplotypes are of special interest in the specific analysis of the male component of DNA mixtures that are very frequently faced by forensic labs in sexual assault cases. Another application of Y-STR haplotyping using a commercial multiplex PCR system of 17 loci, is the identification of human remains and missing persons checked against the appropriate reference samples of paternal relatives. As regards X-chromosome STR analysis, it provides complementary analysis in some kinship deficiency cases.

The automated sequence analysis of mitochondrial DNA hypervariable regions is especially important in the analysis of forensic samples negative for nuclear DNA, or when it is necessary to identify human remains by comparing them with maternal relatives.

Multiplex PCR-CE assays of autosomal SNP markers have been developed for human individualization of degraded samples and as a complementary tool in paternity testing. The analysis of autosomal, Y-chromosome, and mtDNA SNPs with well-differentiated allele or haplotype frequencies among global population-groups has been also applied to the inference of likely ancestry.

DNA Electrophoresis Protocols for Forensic Genetics offers a comprehensive coverage of the most modern current electrophoretic protocols and interpretation guidelines used in forensic genetics for the analysis of amplified human DNA fragments and DNA sequencing.

It includes protocols for profiling of autosomal STRs (see Chapters 1–3, and 13–15), Y-STRs (see Chapter 4), X-STRs (see Chapter 5), autosomal SNPs and INDELS (see Chapters 6–8, 10), Y-SNPs (see Chapter 9), mtDNA-SNPs (see Chapter 11), and mtDNA hypervariable regions HV1 and HV2 (see Chapters 19–21).

Besides the use of DNA electrophoresis on different applications for human identification (criminal investigations, kinship analysis, degraded samples, low template DNA, etc.) the book also covers some interesting electrophoretic protocols for molecular identification of nonhuman species with interest in forensic botany (see Chapters 17 and 18), forensic veterinary (see Chapters 16, 22, and 23), and microbial forensics (see Chapter 26).

Finally the book explores novel forensic applications of capillary electrophoresis to target messenger RNA markers for body fluid identification (see Chapter 12), and some forensic applications of microchip capillary electrophoresis (see Chapters 24 and 25).

DNA Electrophoresis Protocols for Forensic Genetics has been written by highly recognized professionals and specialists from different forensic DNA laboratories around the world dealing with casework and using validated technology and high quality standards. This book was made possible, thanks to them.

Madrid, Spain   *Antonio Alonso*

Contents

PART III   ADVANCES IN MICROCHIP CAPILLARY ELECTROPHORESIS

Contributors

ANTONIO ALONSO • *Instituto Nacional de Toxicología y Ciencias Forenses, Servicio de Biología, Madrid, Spain*

MILENA ALÙ • *Department of Diagnostic and Laboratory Service and Legal Medicine, Section of Legal Medicine, University of Modena Reggio Emilia, Modena, Italy*

CÍNTIA ALVES • *IPATIMUP—Institute of Pathology and Molecular Immunology of the University of Porto, Porto, Portugal*

ANTÓNIO AMORIM • *IPATIMUP—Institute of Pathology and Molecular Immunology of the University of Porto, Porto, Portugal; Faculdade de Ciências da Universidade do Porto, Porto, Portugal*

SYLVAIN AMORY • *International Commission on Missing Persons, Sarajevo, Bosnia and Herzegovina*

BARBARA VAN ASCH • *IPATIMUP—Institute of Pathology and Molecular Immunology of the University of Porto, Porto, Portugal; Faculdade de Ciências da Universidade do Porto, Porto, Portugal*

JACK BALLANTYNE • *National Center for Forensic Science, University of Central Florida, Orlando, FL, USA*

BURKHARD BERGER • *Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria*

CORDULA BERGER • *Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria*

ANA BILIĆ • *International Commission on Missing Persons, Sarajevo, Bosnia and Herzegovina*

CLAUS BØRSTING • *Section of Forensic Genetics, Department of Forensic Medicine, University of Copenhagen, Copenhagen, Denmark*

ZORAN M. BUDIMLIJA • *Department of Forensic Biology, New York City Office of Chief Medical Examiner, New York, NY, USA*

THERESA A. CARAGINE • *Department of Forensic Biology, New York City Office of Chief Medical Examiner, New York, NY, USA*

ANDREA CIAMMARUCONI • *Histology and Molecular Biology Section, Army Medical Research Center, Rome, Italy*

MICHAEL D. COBLE • *Biochemical Science Division, National Institute of Standards and Technology, Gaithersburg, MD, USA*

BEATRICE CORRADINI • *Department of Diagnostic and Laboratory Service and Legal Medicine, Section of Legal Medicine, University of Modena Reggio Emilia, Modena, Italy*

HEATHER MILLER COYLE • *Forensic Science Department, Henry C. Lee College of Criminal Justice and Forensic Science, University of New Haven, West Haven, CT, USA*

CORO FERNÁNDEZ • *Instituto Nacional de Toxicología y Ciencias Forenses, Servicio de Biología, Madrid, Spain*

GIANMARCO FERRI • *Department of Diagnostic and Laboratory Service and Legal Medicine, Section of Legal Medicine, University of Modena Reggio Emilia, Modena, Italy*

MANUEL FONDEVILA • *Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain*

MARIA GEPPERT • *Department of Forensic Genetics, Institute of Legal Medicine and Forensic Sciences, Charité-Universitätsmedizin, Berlin, Germany*

IVA GOMES • *IPATIMUP—Institute of Pathology and Molecular Immunology of the University of Porto, Porto, Portugal*

SUSAN A. GREENSPOON • *Virginia Department of Forensic Science, Richmond, VA, USA*

LEONOR GUSMÃO • *IPATIMUP—Institute of Pathology and Molecular Immunology of the University of Porto, Porto, Portugal*

CORDULA HAAS • *Institute of Legal Medicine, Forensic Genetics, University of Zürich, Zürich, Switzerland*

ERIN HANSON • *National Center for Forensic Science, University of Central Florida, Orlando, FL, USA*

CAROLYN R. HILL • *Division of Biochemical Science, National Institute of Standards and Technology, Gaithersburg, MD, USA*

RENÉ HUEL • *International Commission on Missing Persons, Sarajevo, Bosnia and Herzegovina*

JODI A. IRWIN • *Armed Forces DNA Identification Laboratory (AFDIL), Rockville, MD, USA*

EDIN JASARAGIĆ • *International Commission on Missing Persons, Sarajevo, Bosnia and Herzegovina*

MARIA VICTORIA LAREAU • *Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain*

ADRIAN LINACRE • *School of Biological Sciences, Flinders University, Adelaide, SA, Australia*

PENG LIU • *UCSF/UC Berkeley Joint Graduate Group in Bioengineering, University of California, Berkeley, CA, USA; Sandia National Laboratories, Livermore, CA, USA*

ODILE M. LOREILLE • *Armed Forces DNA Identification Laboratory (AFDIL), Rockville, MD, USA*

JUAN ANTONIO LUQUE • *Servicio de Biología, Instituto Nacional de Toxicología y Ciencias Forenses, Barcelona, Spain*

MANUEL CRESPILLO MARQUEZ • *Instituto Nacional de Toxicología y Ciencias Forenses, Servicio de Biología, Barcelona, Spain*

RICHARD A. MATHIES • *UCSF/UC Berkeley Joint Graduate Group in Bioengineering, University of California, Berkeley, CA, USA; Department of Chemistry, University of California, Berkeley, CA, USA*

BRUCE MCCORD • *Department of Chemistry and Biochemistry, Florida International University, Miami, FL, USA*

MARTA MONTESINO • *Instituto Universitario de Investigación en Ciencias Policiales (IUICP), Comisaría General de Policía Científica, Servicio de Analítica, Laboratorio de ADN, Madrid, Spain*

NIELS MORLING • *Section of Forensic Genetics, Department of Forensic Medicine, University of Copenhagen, Copenhagen, Denmark*

WALTHER PARSON • *Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria*

THOMAS J. PARSONS • *International Commission on Missing Persons, Sarajevo, Bosnia and Herzegovina*

RUI PEREIRA • *IPATIMUP—Institute of Pathology and Molecular Immunology of the University of Porto, Porto, Portugal*

CHRISTOPHER PHILLIPS • *Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain*

LOURDES PRIETO • *Instituto Universitario de Investigación en Ciencias Policiales (IUICP), Comisaría General de Policía Científica, Servicio de Analítica. Laboratorio de ADN, Madrid, Spain*

LUTZ ROEWER • *Department of Forensic Genetics, Institute of Legal Medicine and Forensic Sciences, Charité-Universitätsmedizin, Berlin, Germany*

PAULA SANCHÉZ-DIZ • *Genetic Unit, Department of Pathologic Anatomy and Forensic Sciences, Institute of Legal Medicine, University of Santiago de Compostela, Galicia, Spain*

JAMES R. SCHERER • *Department of Chemistry, University of California, Berkeley, CA, USA*

ROBYN THOMPSON • *Department of Chemistry and Biochemistry, Florida International University, Miami, FL, USA*

CARMEN TOMAS • *Section of Forensic Genetics, Department of Forensic Medicine, University of Copenhagen, Copenhagen, Denmark*

PETER M. VALLONE • *Biochemical Science Division, National Institute of Standards and Technology, Gaithersburg, MD, USA*

STOJKO VIDOVIĆ • *International Commission on Missing Persons, Sarajevo, Bosnia and Herzegovina*

STEPHANIE HI YEUNG • *UCSF/UC Berkeley Joint Graduate Group in Bioengineering, University of California, Berkeley, CA, USA*

SILVIA ZOPPIS • *Laboratory of Forensic Genetics, Department of Anatomy, Histology, Forensic Medicine and Locomotor System, University of Rome "Sapienza", Rome, Italy*

Part I

Current Protocols on Capillary Electrophoresis of Amplified DNA Fragments

Chapter 1

An Overview of DNA Typing Methods for Human Identification: Past, Present, and Future

Robyn Thompson, Silvia Zoppis, and Bruce McCord

Abstract

This chapter presents a brief introduction to the historical development of current technologies used in DNA analysis for human identification. The text describes the development of the PCR and short tandem repeats along with subsequent advances in instrumentation such as real-time PCR and capillary electrophoresis. These techniques have brought about a revolution in DNA typing methods through increased efficiency and the application of multiplex fluorescence detection. More recently the development of new STR based typing methods utilizing mini- and Y-STR PCR multiplexes has increased the flexibility of the investigator, permitting the analysis of inhibited and degraded DNA. Future directions for DNA typing are also discussed, including the development of methods for touch samples based on low copy DNA analysis and the determination of tissue/cell type.

Key words: Capillary electrophoresis, DNA typing, Forensic, PCR, Real-time PCR, Short tandem repeats

1. Introduction: An Historic Perspective of DNA Typing

Deoxyribonucleic acid (DNA) typing is a powerful tool for human identity testing and is well established in the forensic community. It is used to aid in the identification of perpetrators of violent crimes such as murder and rape, in burglary and paternity cases and in identification of the remains of missing persons. More recently, technological advances such as capillary array electrophoresis have become routine in forensic laboratories around the world and standardized protocols have been validated for presumptive and confirmatory testing, DNA extraction, DNA quantitation, and analysis of short tandem repeats (STRs).

Although the existence of variable loci in human DNA has been recognized for a long period of time, it was not until the late 1970s and early 1980s that use of polymorphisms for large-scale mapping of human genes was proposed. Arlene Wyman and Ray White established the foundation for this concept, which used polymorphic DNA loci characterized by variable-length restriction fragments called restriction fragment length polymorphisms or RFLP (1). David Botstein et al. later initiated the use of RFLP in mapping genes of the human genome to indicate genetic differences among individuals (2). However, it was Sir Alec Jeffreys who became the first to describe the scientific technique known today as "DNA fingerprinting" or "DNA typing." Jeffreys, an English geneticist, discovered a set of genetic markers called minisatellites or variable number of tandem repeats (VNTRs), which are highly specific to an individual (3). VNTRs are highly polymorphic regions of DNA composed of a set of discrete repetitive sequences at a defined location in the genome that can be used to differentiate members of a population.

The technique used by Jeffreys to examine VNTRs was RFLP. In RFLP analysis, specific restriction enzymes were used to cut extracted DNA into smaller fragments at defined locations based on 4–6 bp sequence motifs. The fragments were separated by size using gel electrophoresis and then the DNA was transferred to a nylon membrane by the Southern blot technique. A radioactive probe targeting the VNTR sequence of interest was then hybridized to the single-stranded DNA (ssDNA) attached to the nylon membrane for detection. The probed DNA was exposed to X-ray film, producing a "DNA fingerprint." Jeffreys proposed the use of *Hin*fI restriction enzyme digests with multi-locus (oligonucleotide) probes to detect multiple fragments in human DNA (3). The band patterns obtained, while unique to an individual, were difficult to interpret; therefore subsequent methods with single locus probes that hybridize to only one UNTR were used.

The first court cases in which Jeffreys' DNA fingerprint methodology was successfully used occurred in England during the mid 1980s. These cases involved an immigration dispute and a double rape/homicide of teenage girls in 1985 and 1986, respectively (4, 5). The homicide case was prophetic in that it involved the first use of DNA for the purposes of exoneration of an innocent suspect and the first use of DNA databases, issues which are still in dispute today.

The next important milestone in DNA fingerprinting was the development of the Polymerase Chain Reaction (PCR) in 1985 by Kary Mullis. This procedure has lead to significant progress in DNA analysis, especially in the forensic community. The PCR permits the production of multiple "copies" of a defined nucleic acid sequence

in vitro (6). During the PCR reaction, a set of short oligonucleotide sequences known as primers bind to the single stranded DNA (ssDNA) template providing a location for a DNA polymerase to produce a copy of the opposite strand. This procedure is repeated multiple times in a three-step process of DNA denaturation, primer annealing, and extension. First the strands are denatured at 95°C, followed by cooling to a lower temperature (~55°C), which allows the primers to anneal to the single stranded template. The temperature is then raised (~72°C) and a DNA polymerase enzyme extends the primers through incorporation of dinucleotide triphosphates (dNTPs). By doing this, synthesis of the dsDNA target is produced in a rapid cyclic process. In each cycle the PCR generates approximately a twofold increase in the target DNA, theoretically producing millions of copies of template over the course of the process which can involve 28 or more repetitive cycles.

The PCR requires a mixture of reaction components and amplification parameters, which must be optimized to produce sufficient amplification products. The major components, included in a "master mix" which is aliquoted into each PCR tube, are as follows: buffer, dNTPs, polymerase, primers, and magnesium chloride. *Thermus aquaticus* (*Taq*), a species of bacterium found in hot springs, is the source of the thermal stable polymerase used for PCR. The later development of a "hot-start" formulation of *Taq* polymerase, which is inactive until heated to approximately 95°C, was another improvement permitting the amplification of samples without the problems of mispriming and misextension that can occur if the reaction is left too long at room temperature (7). Other PCR components that may be included are non-acetylated bovine serum albumin (BSA) and dimethylsulfoxide, both of which improve the efficiency of the reaction and reduce inhibition.

The PCR quickly became an important analytical method for forensic samples because of its sensitivity, specificity, rapid analysis, and ease of automation. PCR amplification technology permits the analysis of forensic samples with low quantities (<1 ng) of extracted DNA, unlike RFLP methods that require at least 50 ng (8). Additionally, many degraded DNA samples can be amplified and subsequently typed because the alleles are much smaller in size compared to the alleles detected by RFLP analysis (9). PCR provides a valid and reliable approach for the analysis of biological evidence found at crime scenes. Historically, the first PCR-related approach used for forensic purposes was the detection of sequence polymorphisms through the use of allele-specific oligonucleotide (ASO) hybridization probes (10). Approaches involving the detection of PCR amplified length polymorphisms known as minisatellites were also explored. For example, the locus D1S80 consisted of a 16-base sequence that repeated up to 41 times yielding allele sizes from 400 to greater than 800 bp in length (11).

2. Short Tandem Repeat Loci: The Current Methodology

While these new techniques proved to be far more sensitive than the previously used RFLP procedures, problems with mixtures and DNA degradation ultimately led to the use of new markers known as STRs. STR loci differ from VNTRs by having a shorter repeat sequence (2–7 bp) resulting in improved sensitivity and better resistance to degradation. STR typing is the primary method presently used in forensic laboratories. Tetrameric and pentameric repeats are the most widely used type of STR and have been multiplexed with up to 16 loci co-amplified in a single PCR reaction. When compared to other PCR-based methods, STR analysis is faster and reactions can be multiplexed, permitting multiple loci to be amplified in a single run. The STR regions are also highly polymorphic, providing an extremely high power of discrimination (12).

In 1997, the Federal Bureau of Investigation (FBI) Laboratory launched a nationwide effort to establish a set of core STR loci for use within a national DNA database known as CODIS (Combined DNA Index System) (13). The 13 loci selected were CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11. Similar sets of STR markers were also selected in the European Union and elsewhere. In general, a DNA profile obtained using 12 or more STR loci yields a composite genotype frequency of less than 1 in 10^{-15} (12). The incredibly small size of these numbers results from the independent inheritance of the STR alleles, permitting their frequencies in the population to be multiplied together, making DNA typing a very powerful technique.

For convenience, most laboratories use large STR multiplex kits capable of amplifying up to 16 STR loci. Applied Biosystems (Foster City, California) and Promega Corporation (Madison, Wisconsin) are the two of the major vendors of multiplex STR kits. These manufacturers also provide allelic ladders containing mixtures of all common alleles present in each locus in their multiplex kits as standards for accurate genotyping. The kits also contain a locus in the Amelogenin gene that contains a 6 bp deletion in the X chromosome, permitting its use in sex determination as well (14). Overall, these sets of STR loci are more than adequate for addressing most problems in human identification, including paternity testing, missing persons cases, violent crime, and mass disasters. Due to their high level of polymorphism and discrete allele sizes, STR multiplex kits are particularly useful in resolving DNA mixtures.

2.1. DNA Extraction Procedures for STR Typing

To prepare samples for analysis by STR typing, it is generally necessary to collect and extract the DNA from various types of biological evidence. This evidence can include single source samples such as blood or semen stains as well as mixed samples such as rape kits in

which the suspect and victims DNA are comingled. A variety of procedures exist to extract the DNA but all work on the same basic principle: the cell walls are lysed, the protein surrounding the DNA is removed, and the naked DNA is isolated, purified, and quantified. The traditional procedure for DNA extraction is known as organic extraction and involves lysing the cell walls with detergent, reducing disulfide bonds and digesting the proteins with enzymes and DTT, and performing a liquid–liquid extraction using phenol chloroform/isoamyl alcohol to remove lipids and proteins (15). This technique can also be modified to extract mixtures of sperm and vaginal cells by eliminating the DTT in the first step, which digests the vaginal cells, but leaves the strongly cross-linked proteins in the sperm head unchanged. The DNA from the vaginal cells can then be removed by centrifugation, DTT added and the extraction repeated to lyse and recover DNA from the sperm cells. This process is known as a differential extraction.

Alternative extraction procedures using chelex ion exchange resins or sucrose also exist and can sometimes produce better results depending on sample type (16). With the advent of robotic systems, solid phase extraction procedures using silica or ion exchange beads have become important. These techniques usually require a larger quantity of DNA than liquid–liquid extraction, but because the mass of the beads can be controlled, a more reproducible amount of DNA can be delivered, simplifying downstream processing. Typically, the DNA is bound to the bead, washed, and then the pH or salt content of the solution is changed, releasing the purified sample. The washing step can be very useful for badly contaminated samples as it helps to remove PCR inhibitors that often create problems when using organic extraction techniques.

2.2. DNA Quantitation

Once the DNA is isolated, it must be quantified with a human-specific technique. This step is important for several reasons. Firstly, multiplex STR typing is very sensitive to the quantity of input DNA with typical levels of input DNA ranging from 0.5 to 2 ng. Because 16 simultaneous PCR reactions are taking place in each tube, large excesses of DNA can produce imbalanced amplifications. Low levels of sample will also produce poor results. Secondly, because many DNA samples may be comingled with fungal and bacterial DNA, generic methods for quantifying DNA can produce misleading results.

Currently the best procedure for performing quantification, utilizes real-time PCR. Real-time PCR is performed by measuring the rate of amplification of one or more human-specific DNA sequences at an early stage of the PCR process (17). During this time, the quantity of amplified DNA is proportional to the quantity of original template as the relative level of sample and reagents is optimum. Measurements are made using specially designed PCR instruments that can determine the increase in dsDNA using

fluorescent probes or intercalating dyes. The advantage of real-time PCR over other previous developed techniques involving hybridization or fluorescence detection is its sensitivity, wide dynamic range, and compatibility with automation (18, 19).

2.3. Separating STR Alleles by Capillary Electrophoresis

Genetic analyzers, using capillary gel electrophoresis, are the most popular systems for separating STR alleles, Fig. 1. These systems have almost completely supplanted slab gel electrophoresis in forensics due to their ease of automation, high efficiency, and speed (20, 21). Both single capillary and capillary array systems are available. The basic design of the capillary gel electrophoresis system involves the use of one or more narrow 50 μm id capillaries, vials filled with buffer, and sets of electrodes connected to a high-voltage power supply. A robotic autosampler fills the capillary with sieving polymer and utilizes an applied voltage to inject a PCR product that has been denatured through dilution in formamide. The amplified DNA fragments are then separated in a solution of entangled linear polymers dissolved in a TAPS buffer containing urea and other denaturants (22). The polymer solution coats the capillary walls and acts as a sieve to separate ssDNA fragments by size as they move down the capillary under the influence of an applied electric field (23). The smaller-sized DNA amplicons interact less with the sieving polymers and therefore elute first from the column. Individual DNA amplicons, previously modified with various fluorescent dyes at their 5′end, are detected by laser-induced fluorescence near the distal end of the capillary. Different dye labels can be distinguished by through their specific emission spectrum and are visualized using a grating coupled to a charged coupled device (CCD) (8). Complex computer algorithms are then utilized to minimize problems with dye overlap that can produce ghost peaks known as pull-up in uncorrected or overloaded data.
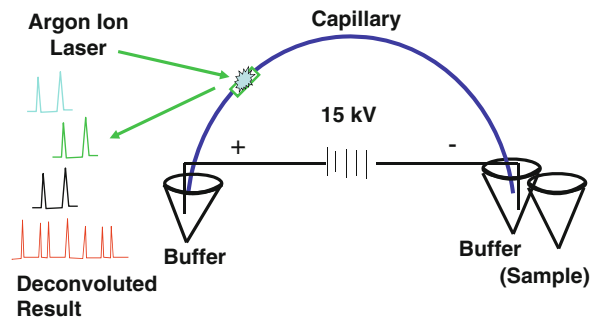
Fig. 1. Schematic of a capillary electrophoresis instrument. The capillary ends and electrodes (connected to a high-voltage power supply) are placed in vials of buffer solution. The sample replaces one of the buffers and is introduced into the capillary. When an electric current is passed across the capillary, analytes are separated based on size and detected by a fluorometric detector. The fluorescent emission is collected and deconvoluted to produce allele peaks.

Modern instruments can simultaneously detect up to five dyes, permitting rapid detection of a wide variety of labeled PCR products. Overall, the loci are distinguished by size and dye label, making the electrophoresis system capable of detecting all 16 genetic loci in a multiplex kit in a single run (24). The data are represented in the form of an electropherogram, where individual peaks represent the different alleles present at each particular locus (Fig. 2).

To maintain sufficient precision to type each DNA fragment to the nearest basepair (+/–0.17 bp), commercial multiplex STR kits utilize allelic ladders and internal sizing standards to permit high-precision typing and comparison of results in spite of the fact that samples may be run over different periods of time, in different laboratories, and on different days. To do this the allelic ladder is sized in comparison to the internal sizing standard and these values are then compared with similar data from the unknown sample. The amplified alleles are then sized and compared to the allelic ladder results in order to determine the genotype (24). This is done through the use of software programs that are compatible with modern genetic analyzers.
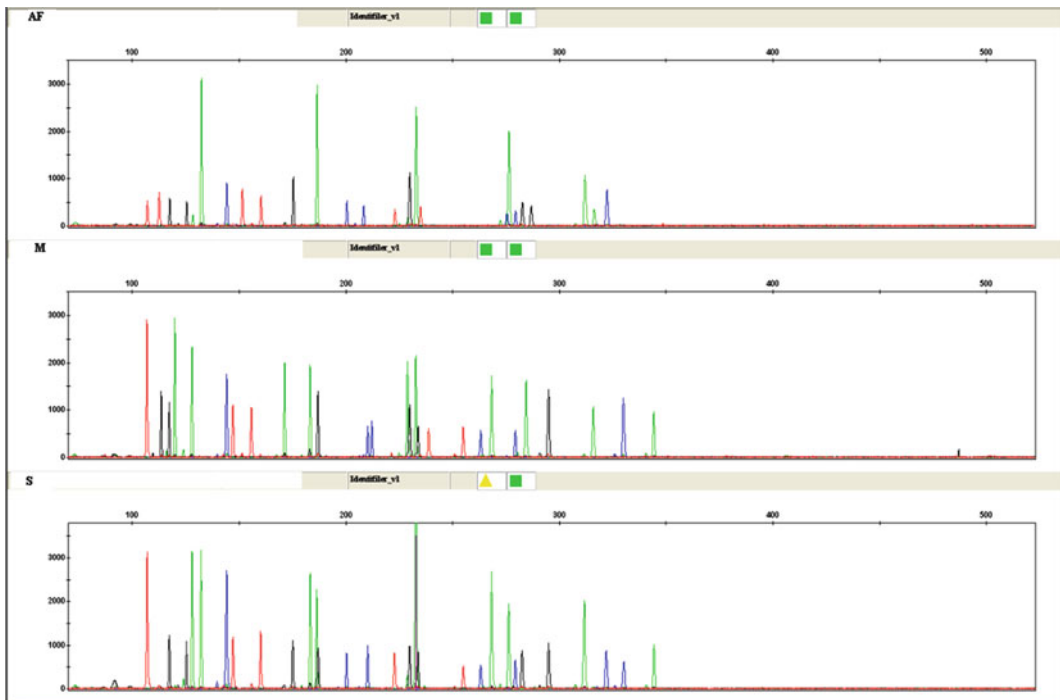
Fig. 2. Electropherogram of the genetic information of an alleged father compared to the mother and son in a paternity examination using AmpFlSTR® Identifiler® PCR Amplification Kit. Note that son shares a single allele with each parent. The colors represent the different fluorescent dyes for each particular locus. The scales at the top of the peaks represent the number of basepairs and the scales to the left represent the intensity of the fluorescent signal in relative fluorescent units (RFUs).

2.4. Improvements in STR Technology

In spite of the advantages associated with STR analysis, there are several drawbacks. A larger number of STRs loci are required due to their lower discrimination power per locus when compared to VNTRs. Interpretation of mixtures is complicated by "stutter," an artifact of the PCR process that produces "false" alleles one repeat shorter than the main allele. Another issue occurs when alleles containing incomplete repeat units appear, resulting in small variations in fragment size. Mutations in the flanking regions of template DNA can result in poor binding of the primer to the DNA. These situations result in poor peak balance or allele dropout at heterozygous loci (8).

Currently the manufacturers continue to upgrade their already validated and highly discriminatory STR systems to tolerate samples that have been highly degraded and inhibited in order to generate profiles that could not be previously amplified. By further reducing the amplicon size, improved recovery of inhibited and degraded samples is possible (21, 25). These reduced-size STR loci can be particularly valuable in the determination of DNA from bone samples, yielding far more useful statistical results than alternative techniques such as mtDNA sequencing (26). Fig. 3 illustrates the application of reduced-size STRs in the analysis of DNA present on a cigarette left behind at a crime scene.
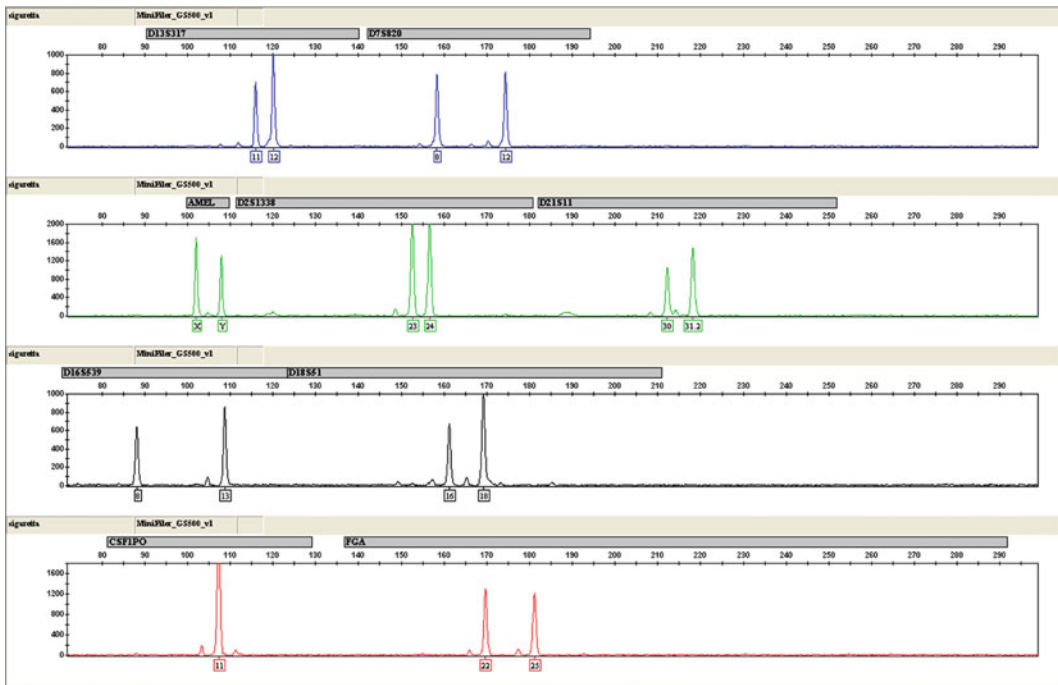
Fig. 3. Analysis of DNA extracted from a cigarette left behind at a crime scene using AmpFlSTR® MiniFiler™ PCR Amplification Kit. Amplicon sizes are reduced in this kit to permit enhanced detection of degraded DNA.

An alternative approach recognizes that much of the problem of recovery of poor quality samples involves the presence of PCR inhibitors (27, 28). To deal with this problem, updated multiplex systems have been designed with improved buffer systems that minimize the impact of PCR inhibition. These next generation systems include Identifiler + (Applied Biosystems and PowerPlex 16 Hot-Start (HS), Promega Corporation, Madison, WI), which can yield profiles from badly inhibited samples as well as improve sensitivity (29, 30). Other recent advances in the field include the development of multiplex kits for typing Y-STRs (31–33) and corresponding multiplex methods for simultaneously detecting and quantifying autosomal and male DNA using real-time PCR (18, 34, 35). Y-STR typing permits the detection of small amounts of DNA in the presence of very large quantities of female DNA. This can be particularly useful in situations involving mixture evidence such as fingernail scrapings, bite marks, and rape kits resulting from aspermatic males. Fig. 4 illustrates the application of Y-STRs in the resolution of a mixture of male and female DNA. The Y-STR loci are separated by size and color and amplified in a manner similar to that used in the autosomal STR kits.
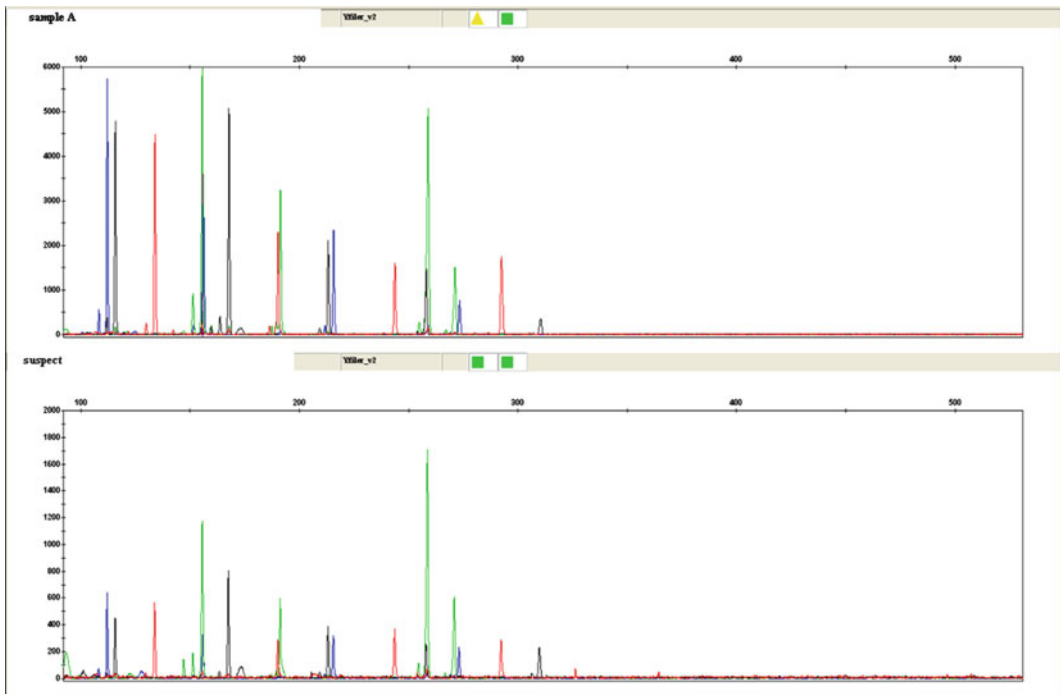
Fig. 4. Comparison between the Y-STR profile obtained from a biological sample at a crime scene (Sample A) and the suspect's profile (AmpFlSTR® Yfiler® PCR Amplification Kit).

3. Future Directions

The improved sensitivity of newly developed STR kits and the development of strategies for amplification of low levels of DNA mean that samples which previously could not or were not analyzed can now produce results. In particular, there is an increasing trend toward the analysis of touch DNA. Such samples often contain very low levels of DNA and can also contain mixtures. This can make detection and interpretation very difficult. To deal with these types of samples, special processing techniques such as post-PCR cleanup, use of additional amplification cycles and extended injection times can result in greatly improved sensitivity in electrophoretic analysis (36). For example, Fig. 5 illustrates the difference between standard DNA analysis of a 65 pg DNA sample and that same sample following dialysis to remove salts. The enhancement in the signal intensity is due to intensification of the electric field during injection that results from the loss of interfering buffer salts during the filtration step.

Unfortunately, these low level samples suffer from stochastic sampling effects which can include peak imbalance, enhanced stutter,
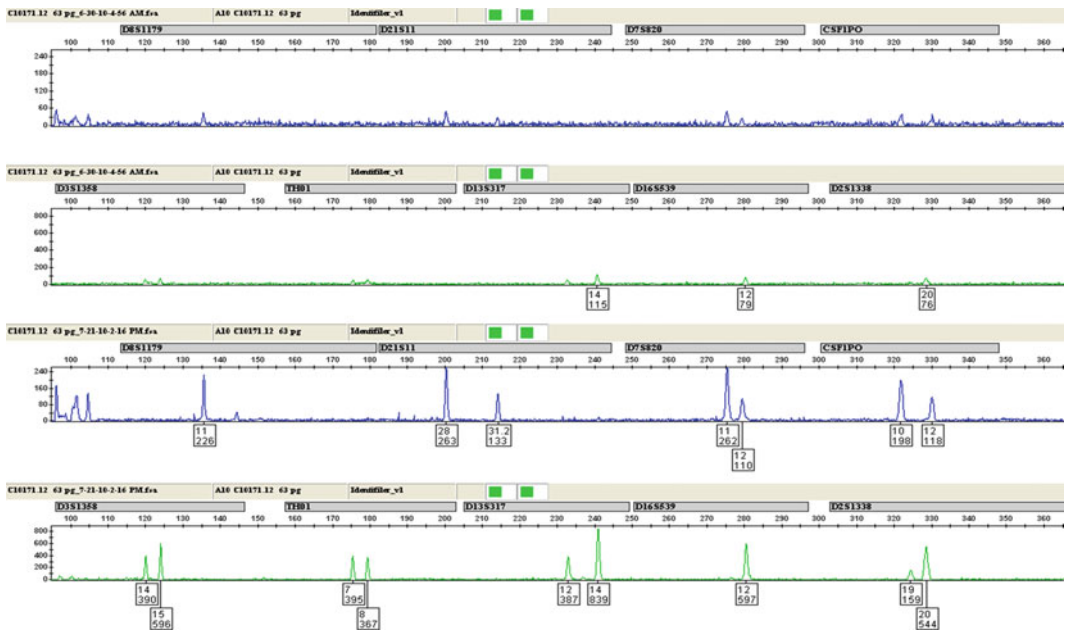
Fig. 5. The effect of post-PCR cleanup on a 65 pg DNA single source DNA sample. The first two panels involve the analysis of a single source sample using the manufacturer's recommended protocols for the AmpFlSTR® Identifiler® PCR Amplification Kit. The second two panels illustrate the detection of that same sample following post-PCR cleanup using spin filtration The enhancement in the signal intensity is due to intensification of the electric field during injection that results from the loss of interfering buffer salts during the filtration step. Note also the imbalance of heterozygous peaks due to stochastic sampling effects.

and allele loss or dropout. (37–39). Several loci in Fig. 5 suffer from an imbalance in peak height. In some cases, these problems can reduce the quality of the results to the point that they become uninterpretable. In addition, because levels of DNA present on surfaces can be so low, competition for amplification with trace levels of DNA deposited on the surface prior to the criminal activity may occur, giving rise to mixtures and unattributable alleles (allele drop-in).

Therefore, in order to properly interpret such samples, strict guidelines have been promulgated that include careful determination of analytical thresholds, the use of replicate analyzes in a profile. Specially designed facilities and the requirement for a minimum number of heterozygous alleles (38, 40). Highly sensitive real-time PCR systems that target multicopy loci such ALU inserts are also required to make precise measurements of low levels of samples (18, 38).

To combat the problem of exogenous DNA, new analytical techniques such as laser microdissection and fluorescence in situ hybridization are being developed, which are capable of identifying, capturing, and amplifying individual cells prior to electrophoresis (41, 42). With such systems, individual sperm cells may be collected eliminating much of the problem with exogenous and female DNA. However, the need for highly experienced personnel and ultraclean facilities will not be diminished by these new techniques. Overall, the challenge with developing guidelines for ultratrace DNA detection is to promulgate clear and universal laboratory practices while recognizing that a multitude of labs exist, each with its own specific protocols and personnel.

Other recently developed areas of research include improved screening of evidence and cell type through RNA expression (43, 44) or epigenetics (45), and the use of genetic markers as probes of physical characteristics such as eye or skin color (46, 47). These techniques help to address the importance of defining the type of DNA found at a crime scene in verifying a witness's story.

4. Conclusions

The development of DNA typing methods by capillary electrophoresis has involved a progression of increasingly sensitive and specific polymorphic markers. The current application using multiplexed STR loci is a powerful tool for forensic identification. Due to their high polymorphism and heterozygosity, STR loci are particularly useful for the resolution of mixed stains, and determination of relatedness of individuals. As we move into the second decade of use of STR typing new applications of these markers are being developed including Y-STR typing, low template analysis, and analysis of highly degraded samples. In addition, auxillary techniques such

as real-time PCR and laser microdissection will further improve our ability to extract information from forensic samples. Regardless of the result of these new and interesting procedures, it is clear that the utility of STR typing by electrophoresis will continue for years to come.

Acknowledgments

References

1. Wyman AR, White R. A highly polymorphic locus in human DNA. *Proc. Nati. Acad. Sci. USA* 1980; 77(11): 6754–6758.

2. Botstein D, White R, Skulnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 1980; 32: 314–31.

3. Jeffreys AJ, Wilson V, Thein SL. Hypervariable 'minisatellite' regions in human DNA. *Nature* 1985; 314: 67–73.

4. Jeffreys AJ, Brookfield JF, Semeonoff R. Positive identification of an immigration test-case using human DNA fingerprints. *Nature* 1985; 317: 818–819.

5. Aronson J. DNA fingerprinting on trial: the dramatic early history of a new forensic technique. *Endeavour* 2005; 29 (3):126–131.

6. Mullis KB. The unusual origin of the polymerase chain reaction. *Sci. Am.* 1990; 56–65.

7. Chou Q, Russell M, Birch DE, Raymond J, Bloch W. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Research* 1992; 20(7): 1717–1723.

8. Butler J, Buel E, Crivelente F, McCord B. Forensic DNA typing by capillary electrophoresis, *Electrophoresis*, 2004; 25(10–11): 1397–1412.

9. Mullis K, Ferr F, Gibbs, R. Polymerase Chain Reaction. Birkhäuser: Boston, MA, 1994.

10. Reynolds R, Sensabaugh G, Blake E. Analysis of genetic markers in forensic DNA samples using the polymerase chain reaction. *Anal Chem.* 1991; 63: 2–15.

11. Duncan GT, Balamurugan K, Budowle B, Smerick J, Tracey ML. Microvariation at the human D1S80 locus. *Int. J. Legal Med.* 1997; 110: 150–154.

12. Tracey M. Short tandem repeat based identification of individuals and parents, *Croat Med. J.* 2001; 42(3): 233–238.

13. Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans and Trinidadians. *J. Forensic Sci.* 1999; 44(6): 1277–86.

14. Sullivan KM, Mannucci A, Kimpton CP, Gill P. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *BioTechniques* 1993; 15: 637–41.

15. Comey CT. DNA extraction strategies of amplified fragment polymorphism analysis *J. Forensic Sci.* 1994; 39: 1254–1269.

16. Vandenberg N, van Oorschot RA, Mitchell RJ. An evaluation of selected DNA extraction strategies for short tandem repeat typing. *Electrophoresis* 1997; 18: 1624–1626.

17. Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR: Real time monitoring of DNA amplification reactions. *Biotechnology* 1993; 11: 1026–1030.

18. Nicklas JA, Buel E. Development of an Alu-based, Real-Time PCR Method for Quantitation of Human DNA in Forensic Samples. *J. Forensic Sci.* 2003; 48(5).

19. Kline MC, Duewer DL, Redman JW, Butler, JM. Results from the NIST 2004 DNA quantification study 2005; 50: 571–578.

20. McCord BR, McClure DM, Jung JM. Capillary Electrophoresis of PCR-Amplified DNA using Fluorescence Detection with an Intercalating Dye *J. Chromat*. 1993; 651: 75–82.

21. Butler JM, Shen Y, McCord BR. The Development of reduced size STR Amplicons as Tools for the Analysis of Degraded DNA. *J. Forensic Sci*. 2003; 48(5): 1054–64.

22. Rosenblum BB, Oaks F, Menchen S, Johnsopn B. Improved single-strand DNA sizing accuracy in capillary electrophoresis. *Nucleic Acids Research* 1997; 25: 3925–3929.

23. Altria KD. Capillary electrophoresis guidebook: Principles, operation, and applications. Humana Press: Totowa, NJ, USA 1995.

24. Moreno LI, McCord B. Separation of DNA for Forensic Applications, *in Handbook of Capillary Electrophoresis*. Landers, J., Ed, CRC Press, Boca Raton, FL. 2008, pp. 761–784.

25. Opel KL, Chung DT, Drábek J, Tatarek NE, Jantz LM, McCord BR. The Application of Miniplex Primer Sets in the Analysis of Degraded DNA from Human Skeletal Remains. *J. Forensic Sci*. 2006; 51(2): 351–356.

26. Parsons TJ, Huel R, Davoren J, Katzmarzyk C, Milos A, Selmanovic A, Smajlovic L, Coble MD, Rizvic A. Application of novel "mini-amplicon" STR multiplexes to high volume casework on degraded skeletal remains. *For. Sci. Int. Gen*et. 2007; 1: 175–179.

27. Bessetti J. An Introduction to PCR Inhibitors. *Profiles in DNA* 2007; 10(1).

28. Opel K, Chung D, McCord BR. A Study of PCR Inhibition Mechanisms Using Real Time PCR. *J. Forensic Sci*. 2010; 55(1): 25–33.

29. Ensenberger MG, Thompson J, Hill B, Homick K, Kearney V, Mayntz-Press KA, Mazur P, McGuckian A, Myers J, Raley K, Raley SG, Rothove R, Wilson J, Wieczorek D, Fulmer PM, Storts DR, Krenke BE. Developmental validation of the PowerPlex 16 HS System: an improved 16-locus fluorescent STR multiplex. *Forensic Sci. Int. Genet*. 2010; 4(4): 257–64.

30. Gefrides L, Donley M, MacMillan K, Klein C, Kahn R. Identifiler plus: validation for forensic casework and implications for quantification of inhibited samples, 2010. http://www.promega.com/geneticidproc/ussymp21proc/abstracts/poster_93.pdf.

31. Sinha SK, Budowle B, Arcot SS, Richey SL, Chakrabor R, Jones MD, Wojtkiewicz PW, Schoenbauer DA, Gross AM, Sinha SK, Shewale JG. Development and validation of a multiplexed Y-chromosome STR genotyping system, Y-PLEX 6, for forensic casework. *J Forensic Sci*. 2003; 48: 93–103.

32. Mulero JJ, Chang CW, Calandro LM, Green RL, Li Y, Johnson CL, Hennessy LK. Development and Validation of the AmpFlSTR® Yfiler™ PCR Amplification Kit: A Male Specific, Single Amplification 17 Y-STR Multiplex System. *J. Forensic Sci*. 2006; 51: 64–75.

33. Krenke B, Viculis L, Richard ML, Prinz M, Milne SC, Ladd C, Gross AM, Gornalld T, Frappier JR, Eisenberg AJ, Barna C, Aranda XG, Adamowicz MS, Budowle B. Validation of a male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex. *Forensic Sci. Int*. 2005; 148: 1–14.

34. Krenke B, Nassif N, Sprecher C, Knox S, Schwandt M, Storts DR. Developmental Validation of a Real-Time PCR Assay for the Simultaneous Quantification of Total Human and Male DNA. *For. Sci. Int. Gen*et. 2008; 3: 14–28.

35. Barbisin M, Fang R, O'Shea CE, Calandro LM, Furtado MR, Shewale JG. Developmental validation of the Quantifiler Duo DNA Quantification kit for simultaneous quantification of total human and human male DNA and detection of PCR inhibitors in biological samples. *J Forensic Sci*. 2009; 54(2): 305–319.

36. Forster L, Thomson J, Kutranov S, Direct comparison of post-28-cycle PCR purification and modified capillary electrophoresis methods with the 34-cycle "low copy number" (LCN) method for analysis of trace forensic DNA samples, *Forensic Sci. Int. Genet*. 2008; 2: 318–328.

37. Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Sci Int*. 2000; 112: 17–40.

38. Caragine T, Mikulasovich R, Tamariz J, Bajda E, Sebestyen J, Baum H, Prinz M. Validation of Testing and Interpretation Protocols for Low Template DNA amples Using AmpFlSTR® Identifiler®, *Croat Med. J*. 2009; 50: 250–267.

39. Budowle B, Eisenberg AJ, van Daal A. Validity of Low Copy Number Typing and Applications to Forensic Science, *Croat. Med. J*. 2009; 50: 207–217.

40. SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories, Scientific Working Group on DNA analysis methods, 2010.

41. Elliot K, Hill DS, Lambert C, Burroughes TR, Gill P. Use of laser microdissection greatly improves the recovery of DNA from sperm on microscope slides. *Forensic Sci Int*. 2003; 137: 28–36.

42. Anslinger K, Bayer B, Mack B, Eisenmenger W. Sex-specific fluorescent labelling of cells for laser microdissection and DNA profiling, *Int. J. Legal Med,* 2007; 121(1): 54–56.

43. Bauer M, Patzelt D. Simultaneous RNA and DNA isolation from blood and semen stains, *Forensic Sci. Int.* 2003; 136: 76–78.

44. Alvarez M, Juusola J, Ballantyne J. An mRNA and DNA co-isolation method for forensic casework samples, *Anal. Biochem.* 2004; 335: 289–298.

45. Frumkin D, Wasserstrom A, Budowle B. Davidson A. DNA methylation-based forensic tissue identification, *Forensic Sci. Int. Genet.* 2011. in press.

46. Liu F, van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, Janssens AC, Kayser M. Eye color and the prediction of complex phenotypes from genotypes. *Curr. Biol.* 2009; 19: R192–193.

47. Valenzuela RK, Henderson MS, Walsh MH, Garrison NA, Kelch JT, Cohen-Barak O, Erickson DT, John Meaney F, Bruce Walsh J, Cheng KC, Ito S, Wakamatsu K, Frudakis T, Thomas M, Brilliant MH. Predicting phenotype from genotype: Normal pigmentation. *J. Forensic Sci.* 2010; 55(2): 315–322.

Website

www.cstl.nist.gov/biotech/strbase

Chapter 2

Capillary Electrophoresis and 5-Channel LIF Detection of a 26plex Autosomal STR Assay for Human Identification

Carolyn R. Hill

Abstract

Multiplex polymerase chain reaction (PCR) is a common method used for DNA typing in forensic and paternity cases. There are numerous commercial short tandem repeat (STR) multiplex assays currently available to the forensic community. These assays amplify the core Combined DNA Index System (CODIS) STR loci for entry into the US. DNA database. Additional non-CODIS loci, which are considered genetically unlinked to the CODIS loci, can be useful in resolving challenging cases such as missing persons and mass disaster victim identification, paternity testing, and immigration testing. An STR multiplex has been successfully developed with 25 non-CODIS autosomal loci plus the sex-typing locus amelogenin for a total of 26 loci in a single 26plex amplification reaction. This chapter will focus on the preparation and the use of the 26plex assay with DNA samples for the purpose of human identification.

Key words: Multiplex PCR, Short tandem repeat, STR marker, Forensic DNA, PCR primers, 26plex, non-CODIS

1. Introduction

Multiplex polymerase chain reaction (PCR) is commonly used for various forensic DNA typing purposes. This type of testing allows for simultaneous amplification of specific target regions of the genome. Currently, there are two commercial short tandem repeat (STR) assays that can amplify 16 genetic loci in one reaction and are widely used in the forensic community. The AmpFlSTR® Identifiler® (Applied Biosystems, Foster City, CA) and PowerPlex® 16 (Promega Corporation, Madison, WI) multiplex kits both include the 13 Combined DNA Index System (CODIS) STR loci that are required by the Federal Bureau of Investigation (FBI) for data entry into the national level of the US DNA database (1, 2).

These 13 STR markers provide a random match probability of approximately 1 in 100 trillion (1, 2). With this level of discrimination, it may not seem necessary to go beyond the CODIS loci and use additional non-CODIS (NC) markers for forensic typing. Indeed, for general forensic matching of an evidence profile to a suspect profile, the 13 CODIS STR loci are currently deemed sufficient. However, additional loci can be beneficial for human identification in scenarios involving relatives such as missing persons/mass disaster victim identification, immigration testing with limited reference samples, deficient parentage testing which is often needed if only one parent and child are tested, and in cases involving incest (3–7). Forensic and kinship testing laboratories are being challenged to answer more difficult genetic questions, and it is important that the necessary tools are in place to address these issues.

We have previously characterized 26 NC reduced-size STR loci (miniSTR) and performed population analyses on US Caucasian, African American, Hispanic, and Asian samples (8, 9). These miniSTRs were designed to span unused chromosomal locations on the 22 autosomes and have been found to be genetically unlinked to the 13 CODIS loci (8, 9). Thus, the single-locus genotype probabilities for the 13 CODIS loci and the 26 NC loci can be multiplied to determine the profile match probability. Originally, assays for the 26 miniSTR markers were developed as panels of three markers (triplexes) with product sizes below 140 base pairs (bp). These assays allowed for the recovery of genetic information from degraded DNA, which is often present in skeletonized missing persons remains or mass disaster victims (3). The loci were arranged into ten miniplexes for further characterization with US population samples.

After the initial testing of the 26 NC markers, it would have been inefficient to perform additional studies with the ten separate miniplexes. Combining all 26 loci into one multiplex reaction would facilitate this task and save resources. The NC markers have moderate-to-high heterozygosity values (>0.5) and narrow allele ranges (5–15 alleles), making them ideal for combining into a large multiplex for further testing (8, 9). The new NC multiplex does not just have utility in our laboratory; it can be beneficial to the forensic community in several ways as well. A larger multiplex can be used for the generation of a high-throughput reference database. Since most of these loci no longer exist as miniSTRs after primer redesign to fit into the multiplex, the 26plex will likely have a limited utility on degraded samples. However, when high quality and quantity DNA samples are available, the 26plex assay could be useful for kinship analysis, complex criminal paternity testing, parentage testing, immigration testing, and in missing persons/mass disaster cases.

A single amplification, five-dye multiplex has been developed to combine 25 of the NC STR loci (see Note 1) plus the sex-typing marker amelogenin (10) in one reaction to enable rapid analysis of reference samples (11, 12). A concordance evaluation was performed with the multiplex to compare genotypes obtained with the previously characterized miniSTR loci to determine null alleles present with the newly designed primer sets (13). The multiplex was also tested with ~400 father/son sample pairs to determine the individual mutation rates of each STR locus (13). Testing on an extended family sample study was performed to examine the potential statistical improvements to adding additional loci to the analysis (13, 14).

To make the information regarding the 26plex available to the forensic community, Standard Reference Material (SRM) 2391b (the PCR-based DNA Profiling Standard) has been updated with certified and reference values for all 26 of the new loci described (15, 16). Bins and panels for genotyping using GeneMapper ID v3.2 (Applied Biosystems) have been designed (17) to allow for a five-dye single amplification multiplex for rapid reference sample typing.

2. Materials

2.1. DNA Samples

The 26plex should be used with human DNA samples at a target amount of 1 ng; however, DNA concentrations as low as 100 pg can be amplified with this multiplex assay. Samples with higher concentrations can be diluted to 1 ng using TE^{-4} buffer (10 mM Tris, 0.1 mM EDTA, pH 8) (see Notes 2 and 3).

2.2. PCR Reagents and Instrumentation

1. 10× GeneAmp® PCR Gold buffer (Applied Biosystems).
2. 25 mM GeneAmp® $MgCl_2$ (Applied Biosystems).
3. 10 mM dNTPs (USB Corporation, Cleveland, OH).
4. 5 U/μL Taq Gold DNA Polymerase (Applied Biosystems).
5. 3.2 mg/mL Bovine Serum Albumin (BSA) (Sigma-Aldrich, St. Louis, MO) (see Note 4).
6. Forward primers are labeled with fluorescent dyes (Applied Biosystems), and reverse primers are unlabeled (Qiagen Operon, Alameda, CA). Primer sequences are listed in Table 1 (see Note 5).
7. TE^{-4} buffer (10 mM Tris, 0.1 mM EDTA, pH 8).
8. Cary Spectrophotometer (Varian, Santa Clara, CA).
9. GeneAmp® 9700 Thermal Cycler (Applied Biosystems).

Table 1
26plex PCR primers used in this assay

Locus name	Forward dye label	Primer sequence (5′–3′)	(Primer), µM
D1GATA113	PET	F - acattaagcacatgcctctttgt	1
		R - **G**atgaactcattggcaaaagga	1
D1S1627	NED	F - catgaggtttgcaaatactatcttaac	2
		R - **G**ttttaattttctccaaatctcca	2
D1S1677	6FAM	F - gcagtcagcttgattgatcc	3
		R - **GTTTCTT**agaatgcaatagcaaatatcagaatg	3
D2S441	NED	F - caaaaggctgtaacaagggcta	1
		R - **G**ttcactctccttcccaaatgtt	1
D2S1776	VIC	F - ttacctgtgagtatgtgtgcgta	1.5
		R - ggtgctaggtgtgctcagga	1.5
D3S3053	VIC	F - tgacacaaatggaccaagaca	2
		R - **GTTTCTT**gagagagcccttgaaatagca	2
D3S4529	NED	F - cccaaaattacttgagccaat	0.75
		R - **G**agacaaaatgaagaaacagacag	0.75
D4S2364	VIC	F - ctaggagatcatgtgggtatgatt	0.75
		R - **G**cagtgaataaatgaacgaatgga	0.75
D4S2408	NED	F - agctgacatcttaccacatgttc	2
		R - **G**tgcttggcatatattaagacactgta	2
D5S2500	VIC	F - gtttactgataaaccaaatgatgtgc	2
		R - **G**taacttaaagggtaaatgtttgcag	2
D6S474	6FAM	F - ggttttccaagagatagaccaatta	1.5
		R - **G**tcctctcataaatccctactcatatc	1.5
D6S1017	NED	F - agatgggaacgatgcagaca	2
		R - gcataaatggatgggtggat	2
D9S1122	VIC	F - gggtatttcaagataactgtagatagg	0.75
		R - **G**cttctgaaagcttctagtttacc	0.75
D9S2157	NED	F - gatcacgccacggta	5
		R - **G**ttctccatttcaaagatcat	5
D10S1248	6FAM	F - cagtaaaaagcaaacctgagca	1
		R - gcttggcaaagagcagatg	1
D10S1435	VIC	F - cacgttgggtttcctgactt	1
		R - **G**cccagctacttgggatgcta	1
D11S4463	6FAM	F - ctgtcccaaggctgagtgtt	3
		R - **GTTTCTT**cgcagggcaaataaaaagaa	3
D12ATA63	6FAM	F - aggtggcagtgagctgtaac	1
		R - **GTT**tcttggattttgagggccta	1

(continued)

Table 1
(continued)

Locus name	Forward dye label	Primer sequence (5′–3′)	(Primer), µM
D14S14343	PET	F - ggctctgatttccaccactg	2
		R - **G**caactcttggaagcccagtc	2
D17S974	NED	F - ggaacacttgagcca	2
		R - gtggactggggtaagg	2
D17S1301	PET	F - aagatgaaattgccatgtaaaaata	2
		R - **G**tgtgtataacaaaattcctatgatgg	2
D18S853	PET	F - acatatataatgtgagaaaggaggagt	2
		R - **G**ttaatgggtgcaacacacc	2
D20S482	PET	F - ctccattctctcacacccaat	1
		R - **G**cacttctggcttttctggttc	1
D20S1082	6FAM	F - acatgtatcccagaacttaaagtaaac	1
		R - **G**cagaagggaaaattgaagctg	1
D22S1045	6FAM	F - ccctgtcctagccttcttatagc	1
		R - **G**ctgtgcccaagttgagagaa	1
Amelogenin	PET	F - ccctttgaagtggtaccagagca	2
		R - gcatgcctaatattttcagggaata	2

The fluorescent dye labels are listed for the forward primers only. The reverse primers are unlabeled. The 5′ guanine (G) residue for "PIGtail" (GTTTCTT) sequence in each reverse primer was added to promote adenylation (3, 18) and is underlined in *bold print*. The approximate primer concentrations are listed in µM (concentration in 20 µL total volume), but must be empirically adjusted for locus-to-locus balance. Amelogenin primers that are used are from Haas-Rochholz et al. (10)

2.3. Capillary Electrophoresis Reagents and Instrumentation

1. 36 cm 3100/3130xl capillary array (Applied Biosystems), 16 capillaries.
2. Matrix Standard SD-33, G5 dye set (Applied Biosystems).
3. Hi-Di™ Formamide (Applied Biosystems).
4. GS500 LIZ internal size standard (Applied Biosystems).
5. A.C.E.™ Sequencing Buffer, 10× Running Buffer (Ameresco, Solon, OH).
6. POP™-4 or POP™-6 Polymer (Applied Biosystems).
7. 16-capillary ABI Prism® 3130xl Genetic Analyzer (Applied Biosystems).

2.4. Data Analysis Software

1. GeneMapper® *ID* v3.2 (Applied Biosystems) (see Note 6).
2. GeneMapper bins and panels (available on STRBase) (see Note 7).

3. Methods

3.1. Primer Set Preparation

1. Fluorescently labeled forward and unlabeled reverse primers are delivered lyophilized and must be reconstituted with TE^{-4} buffer (see Note 8) to the following concentrations:

   Unlabeled reverse primers = 200 μM.

   Labeled forward primers = 100 μM.

2. The unlabeled reverse primers come with specific quantification information listed in pmol and are reconstituted based on this value. Once reconstituted, they are quantified using a UV spectrophotometer absorbance reading at 260 nm. The dye-labeled forward primers do not come with specific quantification information (these primers are ordered at 10,000 pmol), so they are all reconstituted with the same amount of TE^{-4} buffer to a final concentration of 100 μM and are not quantified with a UV spectrophotometer (see Note 9).

3. The target primer mix concentration is 2 μM, but the primer concentrations should be empirically adjusted for locus-to-locus balance (see Note 10). A sample primer mix preparation sheet is included as Table 2. Note that the primer concentrations used in the primer mix calculations are for the reverse primers since those are variable. The forward primers assume a 100 μM concentration for all markers. This example primer mix set-up is for a 100 μL final volume.

3.2. PCR Reagents and Conditions

1. PCR set-up should be performed in a sterile fume hood.

2. Combine the PCR reagents listed in Subheading 2 for final concentrations of the following:

   2 mM $MgCl_2$, 1× PCR buffer, 250 μM dNTPs, 1 Unit Taq Gold, 0.16 mg/mL BSA, and 0.2 μM primer set concentration. Fill remaining volume with deionized water (diH$_2$O).

3. An example of a PCR template is included as Table 3. The final reaction volume in this example is 19 μL Master Mix + 1 μL DNA sample at 1 ng = 20 μL.

4. The PCR amplification conditions using the GeneAmp 9700 are: denaturation for 11 min at 95°C, amplification for 30 cycles of 45 s at 94°C, 2 min at 59°C, and 1 min at 72°C, extension for 60 min at 60°C, and a final soak at 25°C.

3.3. Capillary Electrophoresis on the ABI 3130xl Genetic Analyzer

1. Prior to testing, a 36-cm capillary array must be installed on the 3130xl with POP-4 or POP-6 polymer and 1× A.C.E. buffer (see Note 11).

2. A 5-dye matrix should be run as the G5 spectral calibration under the "G5 filter" with the five dyes of 6FAM, VIC, NED, PET, and LIZ.

Table 2
An example primer mix preparation sheet

Locus	Locus name	Stock reverse concentration (μM)	Reverse singleplex (μL)	Dye-labeled forward (μL)	Water (μL)
1	D1GATA113	143.6	0.70	1	98.30
2	D1S1627	195.9	1.02	2	96.98
3	D1S1677	98.1	3.06	3	93.94
4	D2S441	141.8	0.71	1	98.29
5	D2S1776	137.9	1.09	1.5	97.41
6	D3S3053	103.1	1.94	2	96.06
7	D3S4529	147.4	0.51	0.75	98.74
8	D4S2364	149.0	0.50	0.75	98.75
9	D4S2408	152.4	1.31	2	96.69
10	D5S2500	184.3	1.08	2	96.92
11	D6S474	190.9	0.79	1.5	97.71
12	D6S1017	153.6	1.30	2	96.70
13	D9S1122	184.4	0.41	0.75	98.84
14	D9S2157	79.5	6.29	5	88.71
15	D10S1248	154.2	0.65	1	98.35
16	D10S1435	154.8	0.65	1	98.35
17	D11S4463	161.7	1.86	3	94.64
18	D12ATA63	192.8	0.52	1	98.48
19	D14S1434	181.4	1.10	2	96.90
20	D17S974	78.4	2.55	2	95.45
21	D17S1301	212.5	0.94	2	97.06
22	D18S853	214.2	0.93	2	97.07
23	D20S482	180.0	0.56	1	98.44
24	D20S1082	200.0	0.50	1	98.50
25	D22S1045	146.1	0.68	1	98.32
Amel	Amelogenin	187.8	1.06	2	96.94
			33.21	44.25	22.54

The unlabeled reverse primers are quantified using a UV spectrophotometer absorbance reading at 260 nm. The dye-labeled forward primers are not quantified and are all reconstituted with TE^{-4} to 100 μM. The primer concentrations used in the primer mix calculations are for the reverse primers. The forward primers use 100 μM for all calculations. This example primer mix set-up is for a 100 μL final volume. The target primer mix concentration is ~2 μM, but should be empirically adjusted for locus-to-locus balance

Table 3
An example of a PCR prep sheet for ten reactions

Stock Concentration	26plex	Desired PCR concentration	Volumes to add	Final volumes
	Total volume of reaction =20μl			n=10 reactions
25 mM	Mg concentration	2 mM	1.6 μL	16 μl
2 μM	Primer concentration	0.2 μM	2 μL	20 μl
5 U/μL	Units of Taq	1U	0.2 μL	2 μl
10 mM	dNTP concentration	250 μM	0.5 μL	5 μl
10×	PCR buffer	1×	2 μL	20 μl
3.2 mg/mL	BSA	0.16 mg/mL	1 μL	10 μl
	Water to add		11.7 μL	117 μl
	Master Mix (MM) volume		19 μl	190 μl
	Volume of added template (μL)	1 μl		
	Add 19 μL MM + 1μL sample = 20 μL rxn			

3. Amplification products are diluted in Hi-Di formamide by adding 1 μL PCR product and 0.3 μL GS500 LIZ internal size standard to 8.7 μL of Hi-Di.

4. The samples are run on the 16-capillary 3130xl without prior denaturation of the samples.

5. The samples are injected electrokinetically for 10 s at 3 kV (see Note 12).

6. The STR alleles are then separated at 15 kV at a run temperature of 60°C.

3.4. Analysis Software

1. Data from the ABI 3130xl are exported as .fsa files and are analyzed using GeneMapper ID v3.2 (see Note 6).

2. Bins and panels have been written for use with POP-6 and POP-4 (see Note 7) and are available for download on STRBase (http://www.cstl.nist.gov/biotech/strbase/str26plex. htm#Bins-and-Panels). There are no ladders available for data analysis (see Note 13).

3. The analytical threshold is set to 50 RFU for analysis in GeneMapper ID v3.2 (see Note 14).

4. An example profile for positive control 9947A is illustrated in Fig. 1 (see Note 15).

5. More information about the 26plex assay can be found on STRBase: http://www.cstl.nist.gov/biotech/strbase/str26plex. htm.
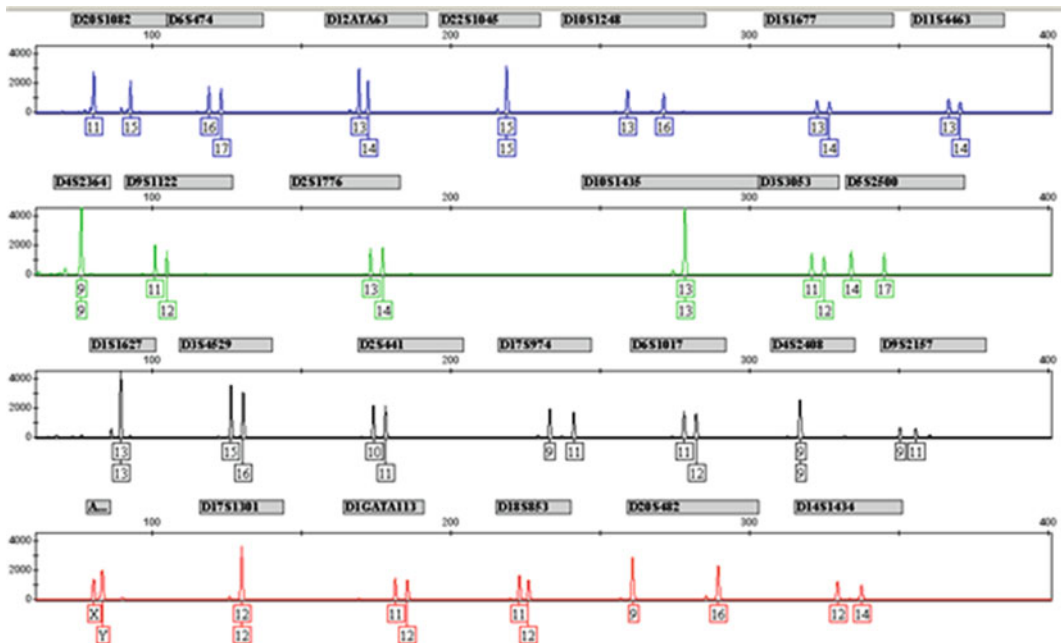
Fig. 1. The final 26plex with 25 autosomal loci plus the sex-typing locus amelogenin for a total of 26 loci. Positive control, 9947A, was used at 1 ng for 30 cycles (12). Primers for the loci were designed so that all of the amplicons present are distributed from 65 basepairs (bp) to less than 400 bp within a five-dye chemistry design with the fifth dye reserved for the sizing standard.

4. Notes

1. In the final 26plex, there were seven loci that used the original miniSTR primer sequences. The primers for the 19 remaining loci were redesigned to fit into the multiplex. Amelogenin was added and the D8S1115 was removed for a total of 26 loci. The D8S1115 was removed due to incompatible primer sequences that consistently caused artifacts to occur in the resulting DNA profiles (12).

2. A sensitivity study was performed (12) to determine the optimal amount of DNA template that can be reliably used with the 26plex. In this study, it was observed that the minimum amount of DNA that can be added without allelic drop-out is 100 pg for 30 cycles. The target amount of input DNA to be added to this assay is 1 ng. Substantial spectral bleed-through between dye channels at several loci was observed with 2 ng of DNA template.

3. At lower amounts of DNA template (<400 pg), there is a consistent artifact peak present within the D4S2364 allele range at ~70 bp (typically between 200 and 450 RFU). The size of the

artifact is proportional to the size of the authentic allele peak heights. However, this artifact peak does not interfere with allele calling. The artifact peak is more apparent with less input DNA as the locus-specific peak heights decrease. The amelogenin peak heights are lower than the allele peaks of other loci with less than 300 pg of template DNA.

4. It is important to use nonacetylated BSA because acetylated BSA interferes in the PCR process.

5. Some of the reverse primers have an additional guanine base (G) or "PIG-tail" sequence (GTTTCTT) added to the 5′ end to help promote full adenylation (3, 18) and are noted in bold and underlined in Table 1.

6. The majority of the 26plex data have been analyzed with GeneMapper ID v3.2. However, GeneMapper ID-X v1.1 and v1.2 (Applied Biosystems) can also be used to analyze the data. The bins and panels were originally written in v3.2, but have been successfully imported into both versions of ID-X.

7. The bins and panels were originally written for samples run on a 36-cm array with POP-6 polymer (12). However, the 26plex has since been run on a 36-cm array with POP-4 polymer. The bins and panels have been adjusted accordingly and both sets are available for download on STRBase. If any other 3130xl parameters are used (i.e., 50-cm array or POP-7 polymer), the bins and panels will have to be further adjusted to fit the data because there are no allelic ladders available for the 26plex at time of publication.

8. It is imperative to use TE^{-4} buffer to reconstitute the lyophilized primers instead of diH$_2$O. It has been our experience that using TE^{-4} buffer reduces the presence of dye artifacts that are present in the final DNA profile compared to diH$_2$O. The dye artifacts can be recognized as being fairly broad peaks that possess the spectrum of one of the dyes used for genotyping (12). If dye artifacts are still present, these can be removed post-PCR using gel filtration cartridges such as Edge Columns (Edge Biosystems, Gaithersburg, MD).

9. The initial primer sequences can be ordered at any concentration. The initial concentrations listed serve as guidelines only. The final concentrations are listed as approximate values due to the fact that it is not possible to initially quantify the fluorescent dye-labeled forward primers.

10. The concentrations of the primer pairs were kept the same when preparing the primer mix. If it was necessary for the concentration of a locus to be adjusted, the forward and reverse primers were adjusted equally. The balancing of the primer mix is an essential step in the 26plex protocol and great care must be made in empirically increasing or decreasing the primers to

generate balanced PCR products (as demonstrated by peak heights).

11. As mentioned in Note 6, POP-6 was originally used for the development of the 26plex. During the development of the 26-plex, we used a single 3130xl instrument for fragment analysis and sequencing. This allowed for higher resolution separations to be achieved while eliminating the need to change the instrument between fragment analysis and sequencing. In addition, a 36-cm array was used, but was installed as a 50-cm array on the instrument. This is because there is no 36 cm, POP-6 module available within the Data Collection Software v3.0 that is installed with the 3130xl instrument. There is, however, a 50 cm, POP-6 module that was used for analysis. Also, as mentioned in Note 6, a 36-cm array with POP-4 can also be used to run the 26plex – this is the recommended human identification (HID) configuration.

12. The injection time and voltage can be adjusted if off scale data and bleed-through between dye channels occur. Typically for a "low injection," the samples were injected for 5 s at 2 kV.

13. As of the time of publication, no allelic ladders are available for the 26plex data analysis. Instead, genotyping and allele sequencing were performed with Standard Reference Material (SRM) 2391b, PCR-based DNA Profiling Standard for all the components including ten genomic DNA samples, and two cell lines (12 samples total) using the 26 additional loci. Certified and reference values were assigned to all resulting alleles and the Certificate of Analysis was updated to include this new information (15). These values can serve as a way to calibrate the genotypes observed with analyzing data from the 26plex (16).

14. The analytical threshold for the 3130xl used at NIST was set at 50 RFU. The analytical threshold and stochastic thresholds should be determined by the individual laboratory according to validation studies performed using their own instruments.

15. The 26plex is a robust STR multiplex assay with balanced heterozygote peak heights and adequate interlocus peak balance. No PCR or fluorescent dye artifacts were present at 1 ng for 30 cycles that could interfere in the correct genotyping of DNA samples.

Acknowledgments

References

1. Butler JM. Genetics and genomics of core STR loci used in human identity testing. J Forensic Sci 2006;51(2):253–65.

2. Budowle B, Moretti TR., Niezgoda SJ., Brown BL. CODIS and PCR-based short tandem repeat loci: law enforcement tools. Proceedings of the Second European Symposium on Human Identification, Innsbruck, Austria, June 1998. Madison, WI: Promega Corporation. 1998; 73–88; http://www.promega.com/geneticidproc/eusymp2proc/17.pdf. Accessed on September 18, 2008.

3. Butler JM, Shen Y, McCord BR. The development of reduced size STR amplicons as tools for analysis of degraded DNA. J Forensic Sci 2003;48(5):1054–64.

4. Wenk RE, Chiafari FA, Gorlin J, Polesky HF. Better tools are needed for parentage and kinship studies. Transfusion 2003;43:979–81.

5. Wenk RE, Gjertson DW, Chiafari FA, Houtz T. The specific power of parentage exclusion in a child's blood relatives. Transfusion 2005;45:440–4.

6. Goodwin W, Ballard D, Simpson K, Thacker C, Syndercombe Court D, Gow J. Case study: paternity testing – when 21 loci are not enough. In: Doutremepuich C, Morling N, editors. Progress in forensic genetics 10. Amsterdam, The Netherlands. International Congress Series. 2004;1261:460–2.

7. Wenk RE. Incest indices from microsatellite genotypes of mother-child pairs. Transfusion 2008;48:341–8.

8. Hill CR, Kline MC, Coble MD, Butler JM. Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples. J Forensic Sci 2008;53(1):73–80.

9. Coble MD, Butler JM. Characterization of new miniSTR loci to aid analysis of degraded DNA. J Forensic Sci 2005;50:43–53.

10. Haas-Rochholz H, Weiler G. Additional primer sets for an amelogenin gene PCR-based DNA-sex test. Int J Legal Med 1997;110:312–5.

11. Butler JM. Constructing STR multiplex assays. Methods in Molecular Biology: Forensic DNA Typing Protocols (Carracedo, A., ed.), Humana Press, 2005;297:53–66.

12. Hill CR, Butler JM, Vallone PM. A 26plex autosomal STR assay to aid human identity testing. J Forensic Sci 2009;54(5):1008–1015.

13. Butler JM, Hill CR, Decker AE, Kline MC, Reid TM, Vallone PM. New autosomal and Y-chromosome STR loci: characterization and potential uses. Proceedings of the Eighteenth International Symposium on Human Identification 2007. See http://www.promega.com/geneticidproc/ussymp18proc/oralpresentations/Butler.pdf.

14. O'Conner KL, Butts ER, Hill CR, Butler JM, Vallone PM. Evaluating the effect of additional forensic loci on likelihood ratio values for complex kinship analysis. Proceedings of the 21st International Symposium on Human Identification 2010.

15. NIST Certificate of Analysis for SRM 2391b: https://srmors.nist.gov/view_cert.

cfm?srm=2391B. Accessed on December 23, 2010.

16. SRM 2391b genotypes for 26 non-CODIS autosomal STR loci: http://www.cstl.nist.gov/biotech/strbase/miniSTR/miniSTR_NC_loci_types.htm. Accessed on December 23, 2010.

17. Additional 26plex data and information: http://www.cstl.nist.gov/biotech/strbase/

str26plex.htm. Accessed on December 23, 2010.

18. Brownstein MJ, Carpten JD, Smith JR. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. BioTechniques 1996;20(6):1004–10.

Chapter 3

Capillary Electrophoresis of MiniSTR Markers to Genotype Highly Degraded DNA Samples

Michael D. Coble

Abstract

The amplification of short tandem repeat (STR) markers throughout the human nuclear DNA genome are used to associate crime scene evidence to the perpetrator's profile in criminal investigations. For highly challenged or compromised materials such as stains exposed to the elements, skeletal remains from missing persons cases, or fragmented and degraded samples from mass disasters, obtaining a full STR profile may be difficult if not impossible. With the introduction of short amplicon STR or "miniSTR" typing, it is possible to obtain STR genetic information from highly challenged samples without the need to sequence the hypervariable regions of the mitochondrial DNA (mtDNA) genome. Non-Combined DNA Index System (CODIS) STR markers have been developed to obtain information beyond the core CODIS loci. This chapter will focus on the steps necessary to prepare and use one of the non-CODIS (NC) multiplexes, NC01 (Coble and Butler 2005), for analysis on capillary electrophoresis instrumentation.

Key words: Mini-STRs, PCR, Degraded DNA, STR, DNA typing

1. Introduction

Short tandem repeat (STR) markers amplified by the polymerase chain reaction (PCR) and separated by capillary electrophoresis have greatly advanced the field of forensic DNA profiling over the last decade. In the United States, a set of 13 common core STR markers (known as CODIS, Combined DNA Index System) has been used for criminal casework and databasing for the National DNA database. Presently, there are two commercially available STR multiplex kits for DNA testing ([1, 2]). Each kit contains all 13 CODIS markers, Amelogenin for sex determination, and two STR markers unique to each provider's kit (D2S1338 and D19S433 from Applied Biosystems (Foster City, CA); Penta D and Penta E from Promega Corporation (Madison, WI)).

For most crime scene samples containing high-quality DNA at a concentration of at least 1.0 ng/µL postextraction, a full STR profile can (usually) easily be obtained with the commercial kits barring the presence of an inhibitor to the PCR process. When the sample quality is challenged due to chemical, enzymatic, or bacterial damage (to note only a few), it is likely that only a partial DNA profile may be obtained. With moderately degraded DNA samples, the high molecular weight markers (>250 bp) are typically absent from the STR profile. For highly degraded samples, it may be possible to only recover a few of the lower molecular weight markers. The net effect of partial DNA profiles resulting from degraded DNA is that the random match probability increases, lowering the significance of a match. For example, if only 4–5 loci are amplified from a degraded sample, the random match probability between the evidence and the suspect (or a victim in an unidentified unknown scenario) may only be in the thousands or tens of thousands.

Many of the highly challenged samples such as those encountered at crime scenes, at mass disasters, or with skeletal remains from missing persons investigations may require mitochondrial DNA (mtDNA) testing to generate a result. Although mtDNA testing can be a useful tool for human identification, the process is costly, time-consuming, susceptible to contamination, and laborious to perform. MtDNA testing can also be limited by a low power of discrimination for common mtDNA types (3). Autosomal SNPs have been proposed as markers for obtaining genetic information from degraded samples (4) since these targets can be made into small amplicons (typically under 100 bp). Limitations of using autosomal SNPs at this time include the lack of a national DNA database populated with SNP profiles and the difficulty to conduct mixture deconvolution with biallelic SNPs (5).

As an alternative to mtDNA sequencing and autosomal SNP testing, reduced amplicon STRs or "miniSTRs" have been developed recently to obtain DNA profiles from challenged samples (6–10). For large STR multiplexes, the goal is to space as many markers as possible into each dye channel. The primers for higher molecular weight amplicons are spaced at a distance away from the repeat (e.g., 400 bp in Fig. 1). MiniSTRs are created by placing the amplification primers as close to the core STR repeat as possible (Fig. 1). By placing the primers adjacent to the core repeats, one can generate a smaller PCR amplicon (e.g., 100 bp in Fig. 1) that has a greater possibility of being amplified from a degraded DNA sample (and also retain the same information (genotype) as the longer-sized amplicon).

MiniSTRs have been critical for mass disaster victim identification efforts such as the identification of remains from the September 11, 2001 attack on the World Trade Center (9) where nearly 20% of the 850 DNA-only identifications were made with mini-STRs

STR primers from a high molecular weight marker
in a standard commercial kit
~400bp

GATAGATAGATA

Mini-STR primers
~100bp
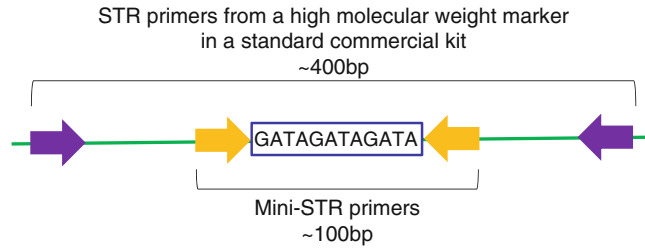
Fig. 1. The concept of a "miniSTR." In commercially available STR kits, each marker is efficiently spaced apart to maximize the number of markers in a particular dye channel. High molecular weight marker primers can amplify regions around 400 bp. In samples that are highly degraded, these primers may fail to amplify enough copies to produce an allelic peak. By moving the primers as close to the core repeating motif as possible, the miniSTR primers can produce a substantially smaller amplicon (e.g., around 100 bp) while maintaining the same number of repeats amplified with the commercial kit primers.

(11, 12). Miniplexes of miniSTR loci can also be useful for sorting highly comingled remains in mass grave identification (13, 14). MiniSTRs have been developed for the CODIS loci (9), non-CODIS markers (15, 16), X-chromosomal STRs (17, 18), and Y-STRs (19). Both commercial vendors of STR kits in the US have miniSTR kits now available: MiniFiler (Applied Biosystems) for degraded samples (20) and PowerPlex S5 (Promega) for use as a low-cost sorting tool (21).

The creation of a novel miniSTR multiplex can be performed in a two-step process consisting of "computer work" followed by "laboratory work" and is outlined in two publications from the human identity group at NIST (16, 22). First, using the computer, one identifies candidate STR markers. One valuable resource is the Marshfield Maps from the Marshfield Clinic Center for Human Genetics (http://research.marshfieldclinic.org/genetics/Genetic Research/compMaps.asp). After selecting a set of loci, the next step is to download the sequence data (using the GenBank number from the Marshfield Maps) from the National Center for Biotechnology Information (NCBI) BLAST-nucleotide website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). It is highly suggested that one confirms the exact chromosomal location of the sequence using the University of California Santa Cruz Genome Browser—Human BLAT search of the sequence (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start). Next, a PCR primer design program such as Primer 3 ((23); http://frodo.wi.mit.edu/primer3/) can be used to identify primer sites as close to the repeat region as possible and thus generate the smallest possible amplicon. Finally, once a set of primers has been identified, it is recommended to use a software program such as AutoDimer ((24); http://www.cstl.nist.gov/strbase/AutoDimerHomepage/AutoDimerProgram Homepage.htm) to avoid nonspecific primer–primer combinations

that may form dimerization products, which can reduce the efficiency of the PCR reaction.

Once the miniSTR primer sets have been ordered, the second stage of the process involves "laboratory work." The first step is to optimize the primer balance – first in singleplex PCR reactions, and then in multiplex PCR reactions. It may be necessary to increase the primer pair concentration of one marker or lower the primer pair concentration of another in order to balance the fluorescent signal in each dye channel. The next step is to test the markers on a set of population samples to assess the number of alleles observed, allele frequency, heterozygosity, and other statistical parameters (16). If the markers are suitable for human identification, then it is necessary to determine the allele sizes (i.e., the number of repeats) observed for each marker.

By selecting homozygous alleles from the population samples, it is possible to determine the number of repeats of the allele by sequencing. A pair of sequencing primers should be selected using the Primer 3 software to sequence the flanking regions of the miniSTR amplicon. For example, if the homozygous allele is 140 bp, then one should sequence about 100 bp upstream and downstream of the amplicon. Thus, a region of about 340–400 bp with the homozygous allele in the "center" of the amplicon is produced. Sequencing the amplicon will determine the number of STR repeats and may also reveal SNPs in the miniSTR primer sequence that create null alleles.

Typically, at least two different alleles should be sequenced. Preferably, one should sequence an allele at the low end of the allele range and an allele at the high end of the range. Once the allele sizes in bp are correlated with the repeat numbers, one can create bins and panels for genotyping. It is not necessary to create allelic ladders. Allelic ladders are laborious to make and have the potential to contaminate the lab with amplified PCR product. As an alternative to allelic ladders, one can determine the number of repeats by sequencing one (or several) positive controls such as 9947A, 007, or the NIST SRM 2391b DNA samples. By correlating the number of repeats in the positive control DNA samples, genotyping can be performed by including the controls with each plate of samples.

The following protocol is meant to guide the user from the stage where the "computer work" ends and PCR primer sequences are identified. The three non-CODIS markers described in Coble and Butler (16) as part of NC01 (D10S1248, D14S1434, and D22S1045) are used as an example (see Fig. 2). Sequencing primers, methods, troubleshooting tips, and additional information can be found in Coble and Butler (16) and Hill et al. (22).
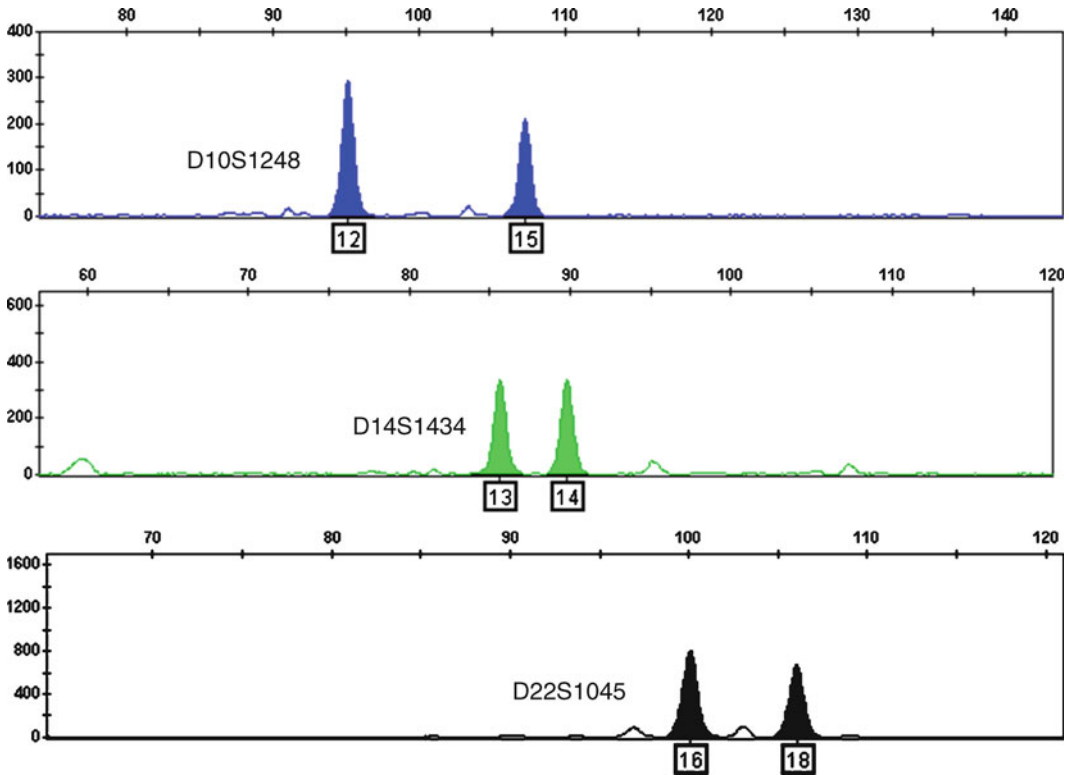
Fig. 2. An example of a sample amplified with the NC01 miniSTR multiplex. Each marker of the NC01 multiplex (D10S1248, *blue channel*; D14S1434; *green channel*; D22S1045; *yellow* (noted as *black*) *channel*) in this sample is heterozygous. Note that the size range of each marker falls under 120 bp, making these markers useful for degraded samples.

2. Materials

2.1. PCR Primers for MiniSTR Multiplex NC01

2.1.1. Ordering of PCR Primers

For convention, all forward primers were synthesized with fluorescent dyes and reverse primers were unlabeled. Labeled forward primers were ordered from Applied Biosystems (Foster City, CA) with the fluorescent dyes 6FAM (blue), VIC (green), and NED (yellow). Unlabeled reverse primers were purchased from Qiagen Operon (Alameda, CA). Primers were ordered lyophilized and desalted.

2.1.2. PCR Primer Sequences

Table 1 contains the labeled and unlabeled primer sequences for the NC01 assay (see Note 1).

2.1.3. Rehydration Buffer

Primers were rehydrated in a solution of Tris Low EDTA buffer (10 mM Tris–HCl and 0.1 mM EDTA), pH = 8.0.

2.2. UV Spectrophotomer

A NanoDrop 2000 (UV spectrophotometer) from Thermo Scientific (Wilmington, DE) was used to determine the concentration of rehydrated primer solutions.

Table 1
PCR primer sequences for the NC01 multiplex amplification

Marker name	Primer type	Prime sequence (5'-3')	Amplicon size (bp)[a]
D10S1248	Forward	(FAM)-TTAATGAATTGAACAAATGAGTGAG	102
	Reverse	**G**CAACTCTGGTTGTATTGTCTTCAT	
D14S1434	Forward	(VIC)-TGTAATAACTCTACGACTGTCTGTCTG	88
	Reverse	**G**AATAGGAGGTGGATGGATGG	
D22S1045	Forward	(NED)-ATTTTCCCCGATGATAGTAGTCT	105
	Reverse	**G**CGAATGTATGATTGGCAATATTTTT	

[a]Based upon the size of the allele present in GenBank

The 5' guanine residue in each reverse primer (underlined and in bold font) was added to promote full adenylation of the amplicons (see Note 1 and ref. (9))

2.3. DNA Polymerase and Buffers for PCR

AmpliTaqGold® DNA polymerase (5 U/μL), 25 mM MgCl$_2$, and 10× GeneAmp® PCR Gold buffer was purchased from Applied Biosystems.

2.4. Deoxy-ribonucleotide Triphosphates

A stock solution containing 10 mM of each deoxyribonucleotide triphosphate (dNTP) was purchased from USB Corporation (Cleveland, OH) (see Note 2).

2.5. Bovine Serum Albumin

Bovine serum albumin (BSA) fraction V lyophilized powder, molecular biology grade, nonacetylated was purchased from Sigma-Aldrich (St. Louis, MO). A stock solution of 3.2 mg/mL was prepared in nanopure water.

2.6. Thermal Cycler

A GeneAmp® PCR System 9700 (Applied Biosystems) was used to amplify the NC01 miniSTR multiplex. The cycler was run in 9600 emulation mode with a ramp rate of 1°C/s.

2.7. Capillary Electrophoresis

Either the ABI 310 or 3130xl Genetic Analyzers (Applied Biosystems) can be used for fragment separation. A 5-dye matrix should be established under the "G5 filter" with the dyes 6FAM, VIC, NED, PET, and LIZ using matrix standard set DS-33 (part number 4318159 for the 310; part number 4345833 for the 3130 from Applied Biosystems). A 36-cm array was used for fragment separation.

2.8. Separation Polymer

Either performance optimized polymer POP-4 or POP-6 (Applied Biosystems) was used for fragment separation (see Note 3).

2.9. Electrophoresis Buffer

ACE 10× sequencing buffer commercially available from Amresco (Solon, OH) was used in a 1× concentration (diluted with dH$_2$O).

2.10. Formamide	PCR amplicons were diluted in Hi-Di formamide purchased from Applied Biosystems prior to electrophoresis.
2.11. DNA Sizing Standard	GeneScan™ 500 LIZ Size Standard was purchased from Applied Biosystems.
2.12. Software Fragment Analysis	The commercial software package GeneMapper ID v3.2 (Applied Biosystems) was used for data analysis. Bins and panels for NC01 are available on STRBase (see Subheading 3.7). There are no allelic ladders available for NC01.

3. Methods

3.1. Primers	Stock primer solutions were reconstituted at 100 µM for labeled primers and 200 µM for unlabeled primers. Primers were rehydrated with appropriate volumes of Tris Low EDTA buffer (see Note 4).
3.1.1. Primer Rehydration	
3.1.2. Quantification of Primers	Unlabeled reverse primers were synthesized with specific concentration information (in pmol) and were reconstituted from this value. Postrehydration, the primers were quantified using a UV Spectrophotometer by determining the absorbance reading at 260 nm. Forward primers do not come with specific concentration values. Labeled primers can be ordered at various scales depending on the amount of primer desired. Typically, for the first time a primer is tested, the smallest scale is ordered (10,000 pmol). Therefore, the labeled primers were all reconstituted with the same amount of Tris Low EDTA buffer to a final concentration of 100 µM and were not quantified by UV spectrophotometer.
3.1.3. Multiplex Primer Mix	The multiplex primer mix concentration was 2 µM for each primer pair. Initially, the concentration of each primer pair in the primer mix was equivalent (see Note 5). As an example, a 100 µL final volume primer mix for the three NC01 markers (D10S1248, D14S1434, and D22S1045) would be created by taking 1.0 µL of each labeled forward primer (200 µM solution) plus 0.5 µL of each unlabeled reverse primer (total volume = 4.5 µL) plus 95.5 µL of dH_2O (see Note 6).
3.2. PCR Amplification Conditions	PCR amplification of the NC01 multiplex can be performed in a total volume of 10 µL for high-quality samples or 25 µL for degraded DNA samples (see Note 7). Final PCR reagent concentrations were: 1 U of AmpliTaqGold® DNA polymerase (degraded or inhibited DNA samples are typically amplified with 2 U of polymerase), 1× Taq Gold PCR buffer, 250 µM dNTPs, 5 mM $MgCl_2$, 0.16 mg/mL BSA fraction V (this is optional for high-quality samples), and 0.2 µM of each amplification primer pair (see Note 8).

Table 2
PCR preparation worksheet

	10 µL PCR	25µL PCR
	1× Reaction (µL)	1× Reaction (µL)
PCR buffer	$(n+\times)^a \times 1.0$	$(n+\times)^a \times 2.5$
MgCl$_2$	$(n+\times) \times 0.8$	$(n+\times) \times 2.0$
Primer mix	$(n+\times) \times 2.0$	$(n+\times) \times 5.0$
dNTPs	$(n+\times) \times 0.25$	$(n+\times) \times 0.625$
BSA[b]	$(n+\times) \times 0.5$	$(n+\times) \times 1.25$
TaqGold[c]	$(n+\times) \times 0.2$	$(n+\times) \times 1.0$
dH$_2$O	$(n+\times) \times 4.25$	$(n+\times) \times 8.125$
Master Mix vol. + template (µL)	$(n+\times) \times 9.0$ 1.0 per sample	$(n+\times) \times 20.0$ 5.0 per sample

For high-quality genomic DNA, 10 µL reactions can be used to conserve reagents. For degraded samples, 25 µL reactions can be used to increase the amount of DNA in the reaction by adjusting the dH$_2$O content

[a]The $(n+\times)$ refers to the total number of samples (n) plus ($\times$) additional samples for pipetting error. Typically, one additional sample is added for every 12 samples tested. For example, with 96 samples, the $(n+\times) = (96+8) = 104$ reactions

[b]BSA is optional for high-quality samples. If omitted, the amount of water can be increased to 4.75 µL

[c]For low template and challenged samples (25 µL reactions), the amount of TaqGold can be increased to 2 U/µL

For high-quality samples, 1.0 µL of DNA (0.5 ng/µL) is added to the reaction. For degraded DNA samples, 5.0 µL of the total 25 µL reaction is available for DNA (see Note 9). An example of a PCR preparation protocol is found in Table 2.

3.3. Thermal Cycling Conditions

Amplification of the multiplex PCR reaction was performed in the GeneAmp 9700 thermal cycler using the following conditions:

3.3.1. PCR Thermal Cycling Conditions

- 95°C for 10 min
  - 94°C for 1 min
  - 55°C for 1 min
  - 72°C for 1 min
  - (28–34 cycles of PCR)
- 60°C for 45 min
- 25°C soak

For high-quality samples, amplification of 28 cycles was performed. For low template and challenged samples, up to 34 cycles can be amplified (see Note 10).

3.4. Capillary Electrophoresis Conditions

The CE instrument (310 or 3130xl) with a 36 cm capillary array filled with either denaturing POP-4 or POP-6 polymer was utilized for DNA fragment separation. ACE buffer at 1× concentration was used as the running buffer for the CE instrument. The standard run module parameters were: Run temperature = 60°C, injection voltage = 3 kV, injection time = 10 s, run voltage = 15 kV, and run time = 1,800 s (see Note 11).

3.5. Preparation of the Samples for CE Separation and Detection

Samples were prepared for the ABI 3130 with 15-µL Hi-Di formamide (Applied Biosystems), 0.35-µL GS500 LIZ size standard (Applied Biosystems), and 1-µL PCR product. For samples separated on the ABI 310, 19-µL Hi-Di formamide, 0.75-µL GS500 LIZ size standard, and 1-µL PCR product were used. Samples were prepared without prior denaturation and "snap cooling" of the PCR product.

3.6. Allelic Ladders

For the commercially available STR typing kits, allelic ladders for each marker are included in the purchase of the kit and used in the automated genotyping of samples. For the non-CODIS minSTRs, NIST has genotyped the alleles present among commonly used positive controls and genotyped and certified each component of SRM 2391b (http://www.cstl.nist.gov/biotech/strbase/miniSTR/miniSTR_NC_loci_types.htm). It is recommended that these controls be used for correctly genotyping population or casework samples rather than relying upon an allelic ladder.

3.7. Data Analysis

Data files from the CE were exported as .fsa files and imported into GeneMapper*ID* v3.2 (Applied Biosystems). Bins and panels for all of the non-CODIS miniSTRs have been developed for use with POP-6 and are available on STRBase (http://www.cstl.nist.gov/biotech/strbase/miniSTR/miniSTR_Panels_NC_bins_bins.txt) and (http://www.cstl.nist.gov/biotech/strbase/miniSTR/miniSTR_Panels_Panels.txt).

4. Notes

1. Reverse primers lacking a guanine base (G) at the 5′ end of the primer have an additional guanine base (G) or "pig-tail" sequence (GTTTCTT) added to the 5′ end to promote full adenylation (see ref. (9) for more details) and are noted in bold and underlined in Table 1.

2. Stock solutions of dNTPs (in about 1.5 mL) can be aliquoted into smaller volumes (e.g., 50 µL) and stored at –20°C. For each new PCR mastermix, a fresh aliquot can be used to reduce the number of freeze–thaw cycles of the dNTPs.

3. In addition to POP-6, POP-4 and POP-7 can also be used as a separation medium. Depending on the polymer used, slight differences in the amplicon mobility may be observed. The bins and panels in the GeneMapper software should also be changed to reflect any mobility change.

4. Primers rehydrated in Tris Low EDTA buffer rather than in distilled water have better stability and fewer artifacts/dye blobs when stored at 4°C up to 18 months.

5. Not all fluorescent dyes are equivalent. For example, FAM tends to fluoresce more strongly than PET dyes. Therefore, it may be necessary to adjust the primer concentrations to achieve balance among each dye channel. This is more of a qualitative exercise (using peak heights in each dye channel as a guide) that takes trial and error to achieve.

6. It is recommended that primer mixes be stored out of the light and at 4°C instead of the freezer to avoid freeze–thaw cycles. As labeled primers and primer mixes age, the intensity of dye artifacts in the electropherogram increases.

7. The increased volume (25 μL reactions) for degraded/inhibited samples can help to dilute inhibitors.

8. It is important to use nonacetylated BSA rather than acetylated BSA since the latter can cause inhibition of the PCR process (25).

9. For low quantities of DNA, up to 8.125 μL of water in the PCR reaction can be substituted for template, giving a maximum input of 13.125 μL of DNA added to the PCR.

10. It should be noted that enhanced techniques, such as increased polymerase and extra PCR cycles, can give rise to exaggerated stochastic effects with low-template DNA. Proper interpretation guidelines (26), including replicate analyses of the data (26, 27), should be utilized to generate reliable results.

11. Since the alleles of the NC01 miniSTR multiplex are all under 150 bp, it is possible to shorten the run time module for data collection since the default run time collects data for amplicons up to 350–400+ bp fragments. Adjusting the data collection time can decrease the overall run time and increase genotyping throughput.

Acknowledgments/Disclaimer

O'Connor and Erica Butts at NIST are appreciated for their editorial comments. I am also grateful to Dr. Antonio Alonso for the opportunity. Finally, I would like to acknowledge the National Institute of Justice for their funding to our NIST Group through the NIST Office of Law Enforcement Standards interagency agreement 2008-DN-R-121.

The opinions and assertions contained herein are solely those of the author and are not to be construed as official or as views of the National Institute of Justice. Commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Justice, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

References

1. Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA (2004) Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFlSTR Identifiler PCR Amplification Kit. *J Forensic Sci.* **49(6)**: 1265–1277.

2. Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, Tomsey CS, Zachetti JM, Masibay A, Rabbach DR, Amiott EA, Sprecher CJ (2002) Validation of a 16-locus fluorescent multiplex system. *J Forensic Sci.* **47(4)**: 773–785.

3. Coble MD, Just RS, O'Callaghan JE, Letmanyi IH, Peterson CT, Irwin JA, Parsons TJ (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.

4. Dixon LA, Murray CM, Archer EJ, Dobbins AE, Koumi P, Gill P (2005) Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. *Forensic Sci Int.* **154(1)**: 62–77.

5. Gill P, Werrett DJ, Budowle B, Guerrieri R (2004) An assessment of whether SNPs will replace STRs in national DNA databases-Joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM). *Science&Justice* **44(1)**: 51–53.

6. Hellmann A, Rohleder U, Schmitter H, Wittig M (2001) STR typing of human telogen hairs – a new approach, *Int. J. Legal Med.* **114(4–5)**: 269–273.

7. Wiegand P and Kleiber M (2001) Less is more – length reduction of STR amplicons using redesigned primers. *Int. J. Legal Med.* **114 (4–5)**: 285–287.

8. Tsukada K, Takayanagi K, Asamura H, Ota M, Fukushima H (2002) Multiplex short tandem repeat typing in degraded samples using newly designed primers for the TH01,CSF1PO,and vWA loci. *Legal Medicine (Tokyo)* **4**: 239–245.

9. 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(5)**: 1054–1064.

10. Opel KL, Chung DT, Drabek J, Tatarek NE, Jantz LM, McCord BR (2006) The application of miniplex primer sets in the analysis of degraded DNA from human skeletal remains, *J Forensic Sci.* **51(2)**: 351–356.

11. Holland MM, Cave CA, Holland CA, Bille TW (2003) Development of a quality, high throughput DNA analysis procedure for skeletal samples to assist with the identification of victims from the World Trade Center attacks. *Croat Med J.* **44**: 264–272.

12. Biesecker LG, Bailey-Wilson JE, Ballantyne J, Baum H, Bieber FR, Brenner C, Budowle B, Butler JM, Carmody G, Conneally PM, Duceman B, Eisenberg A, Forman L, Kidd KK, Leclair B, Niezgoda S, Parsons TJ, Pugh E, Shaler R, Sherry ST, Sozer A, Walsh A (2005) DNA Identifications after the 9/11 World Trade Center attack. *Science* **310(5751)**: 1122–1123.

13. Parsons TJ, Huel R, Davoren J, Katzmarzyk C, Milos A, Selmanovic A, Smajlovic L, Coble

MD, Rizvic A (2007) Application of novel "mini-amplicon" STR multiplexes to high volume casework on degraded skeletal remains, *Forensic Sci Int Genet.* **1**: 175–179.

14. Marjanović D, Durmić-Pasić A, Kovacević L, Avdić J, Dzehverović M, Haverić S, Ramić J, Kalamujić B, Lukić Bilela L, Skaro V, Projić P, Bajrović K, Drobnic K, Davoren J, Primorac D (2009) Identification of skeletal remains of Communist Armed Forces victims during and after World War II: combined Y-chromosome (STR) and MiniSTR approach. *Croat Med J.* **50(3)**: 296–304.

15. Ohtaki H, Yamamoto T, Yoshimoto T, Uchihi R, Ooshima C, Katsumata Y, Tokunaga K (2002) A powerful, novel, multiplex typing system for six short tandem repeat loci and the allele frequency distributions in two Japanese regional populations. *Electrophoresis* **23(19)**: 3332–3340.

16. Coble MD and Butler JM (2005) Characterization of new miniSTR loci to aid analysis of degraded DNA. *J Forensic Sci.* **50(1)**: 43–53.

17. Asamura H, Sakai H, Kobayashi K, Ota M, Fukushima H (2006) MiniX-STR multiplex system population study in Japan and application to degraded DNA analysis. *Int. J. Legal Med.* **120(3)**: 174–181.

18. Diegoli TM and Coble MD (2011) Development and characterization of two mini-X chromosomal short tandem repeat multiplexes. *Forensic Sci Int Genet. in press.*

19. Park MJ, Lee HY, Chung U, Kang SC, Shin KJ (2007) Y-STR analysis of degraded DNA using reduced-size amplicons. *Int. J. Legal Med.* **121(2)**: 152–157.

20. Mulero JJ, Chang CW, Lagacé RE, Wang DY, Bas JL, McMahon TP, Hennessy LK (2008) Development and validation of the AmpFlSTR MiniFiler PCR Amplification Kit: a MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA. *J Forensic Sci.* **53(4)**: 838–852.

21. Poetsch M, Kamphausen T, Bajanowski T, Schwark T, von Wurmb-Schwark N (2011) Powerplex® ES versus Powerplex® S5--casework testing of the new screening kit. *Forensic Sci Int Genet.* **5(1)**: 57–63.

22. Hill CR, Kline MC, Coble MD, Butler JM (2008) Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples. *J Forensic Sci.* **53(1)**: 73–80.

23. Rozen S and Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Humana Press, Totowa, NJ, pp 365–386. Source code available at http://fokker.wi.mit.edu/primer3/.

24. Vallone PM, Butler JM (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques. **37(2)**: 226–231.

25. Bell PA, Chaturvedi S, Gelfand CA, Huang CY, Kochersperger M, Kopla R, Modica F, Pohl M, Varde S, Zhao RB, Zhao XJ, Boyce-Jacino MT (2002) SNPstream UHT: Ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *Biotechniques Suppl.* pp 70–77.

26. 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(1)**: 17–40.

27. Benschop CC, van der Beek CP, Meiland HC, van Gorp AG, Westen AA, Sijen T (2011) Low template STR typing: Effect of replicate number and consensus method on genotyping reliability and DNA database search results. *Forensic Sci Int Genet. in press.*

Chapter 4

Interpretation Guidelines of a Standard Y-chromosome STR 17-plex PCR-CE Assay for Crime Casework

Lutz Roewer and Maria Geppert

Abstract

Y-STR analysis is an invaluable tool to examine evidence in sexual assault cases and in other forensic casework. Unambiguous detection of the male component in DNA mixtures with a high female background is still the main field of application of forensic Y-STR haplotyping. In the last years, powerful technologies including a 17-locus multiplex PCR assay have been introduced in the forensic laboratories. At the same time, statistical methods have been developed and adapted for interpretation of a nonrecombining, linear marker as the Y-chromosome which shows a strongly clustered geographical distribution due to the linear inheritance and the patrilocality of ancestral groups. Large population databases, namely the Y-STR Haplotype Reference Database (YHRD), have been established to assess the evidentiary value of Y-STR matches by means of frequency estimation methods (counting and extrapolation).

Key words: Y-STR analysis, Sexual assault, Mixed stains, Haplotype frequency, Match probability, Mixture analysis, Population database

1. Introduction

Sexual assault cases represent a significant percentage of cases examined by forensic laboratories. A notable portion of physical evidence submitted in such cases consists of mixtures of body fluids from the victim and the assailant(s), often containing small quantities of DNA which might be highly degraded. Almost every DNA admixture is characterized by unbalanced proportions of a large female (victim's) and a very small male (suspect's) component. While the standard autosomal STR analysis doesn't detect the male component in most of such cases, the Y-STR testing often provides probative results on the suspect. The Y-STR analysis is especially helpful when there is more than one male responsible for the assault, because the mixed male/male pattern can be resolved and a likelihood of donorship

for each of the assailants can be calculated. Y-STR haplotype analysis is also used for nonsexual assault cases where mixed samples are collected from evidence. Performing Y-STR testing can help to identify all males who have contributed to the evidence. During the last years, the advent of highly sensitive Y-STR kits has prompted many forensic genetic laboratories to evaluate and modify their existing workflow for sexual assault and other casework samples where mixtures can be expected (e.g., vaginal swabs, fingernail scrapings, contact stains) aiming to produce additional probative results (1, 2). The detailed protocols to amplify Y-STRs in multiplex PCR reactions followed by capillary electrophoresis are analogous to autosomal STR analysis and provided by the manufacturers of forensic Y-STR kits. This article addresses the issue of interpretation and reporting of Y-chromosome data. Since all Y-chromosomal STRs constitute a single haplotype, a match between profiles can only be evaluated on the basis of the haplotype frequency instead of the genotype frequency (3). Multiplying the individual allele frequencies is not valid as for independently inherited autosomal STRs. Instead, the frequency of the full haplotype has to be assessed in the relevant population. To determine the match probability, a counting and/or an extrapolation ("frequency surveying") method can be applied (4–6). For both methods, large, representative, and quality-assessed databases of haplotypes sampled in appropriate reference populations are required. In this chapter, we will explain how to utilize the Y-chromosome haplotype reference database (YHRD), the largest curated forensic repository of multiregional population samples, for estimating the haplotype frequencies to be used in the decision-making process (7). This database does not work like an offender database as CODIS (USA), FSS (UK), DAD (Germany), or other. This is a population database and it is intended for use in estimating Y-STR haplotype population frequencies for forensic case work purposes. It cannot be used to identify a particular individual whose sample is in the database. Instead, the database is a searchable listing of anonymous 9- to 17-locus haplotypes of unrelated probands directly submitted by forensic, quality-assessed laboratories from around the world and mandatorily published in peer-reviewed journals (8, 9). Each haplotype has a population identifier (geographical and ethnical) and thus it can be assigned to a larger set of samples of common origin and shared demographic history. This set of samples is called a metapopulation (MP) (7). Because between populations exist large differences in Y-chromosome haplotype proportions (resulting in the highest F_{st} values of all forensic markers), a search in predefined metapopulation samples results in frequencies which are realistic estimates of the probability of occurrence and are less affected by population substructuring. It needs to be emphasized, that—due to the nonrecombining nature of Y-chromosomes—even a highly discriminating panel of 17 Y specific STRs is not individual-specific but lineage-specific and a match concerns all men sharing patrilineal descent. This has

to be properly addressed in the match report. In contrast to a match, an exclusion of Y-STR profiles is unambiguous and the exclusion chance increases with the number of Y-STRs employed. Finally, we will address the issue of male–male mixtures and its statistical interpretation using a likelihood ratio calculation (10).

2. Material

1. AmpFlSTR® Yfiler Plus™ PCR Amplification Kit.
2. Y-Chromosome Haplotype Reference Database (www.yhrd. org and tools therein).

3. Methods

3.1. Frequency Estimation

Figure 1a shows a typical example where a standard autosomal analysis with the AmpFlSTR® SEfiler Plus™ PCR Amplification Kit was performed in a crime investigation. A clear female profile was determined at the evidence (a condom). The occurrence of a small Y peak at the Amelogenin locus may indicate a minor male DNA contribution, but only the females' STR signals could be recorded unequivocally. However, the amplification of the same DNA extract with the AmpFlSTR® Yfiler Plus™ PCR Amplification Kit generates a complete, single-source Y haplotype of the unknown suspect (Fig. 1b, Table 1). The high excess of female DNA doesn't influence the outcome of the Y-STR analysis (see Note 4).

To estimate the frequency of this haplotype, the online YHRD was approached (www.yhrd.org/Search/Haplotypes) (see Note 1). The YHRD can be searched for all single alleles and all allele combinations up to a maximum of 17 loci. Please note that the database size differs for different standard haplotype formats because reference samples have been submitted in a 9-locus (minimal haplotype) (11), 12-locus (PowerPlex® Y kit, Promega), or 17-locus format (AmpFlSTR® Yfiler Plus™ PCR Amplification Kit, Applied Biosystems). Release 34 of the YHRD from November 19, 2010 was used here which includes 87,440 9-locus minimal haplotypes (182 populations), 39,349 12-locus haplotypes, and 27,532 17-locus haplotypes (the latter are subsets of the former less-specific haplotypes) (see Note 2).

3.1.1. Instruction

Database Query

1. To search for a haplotype, enter the alleles of interest for each marker by using the drop-down menu. Enter any of the following value types:

   • Standard ladder allele such as "14."
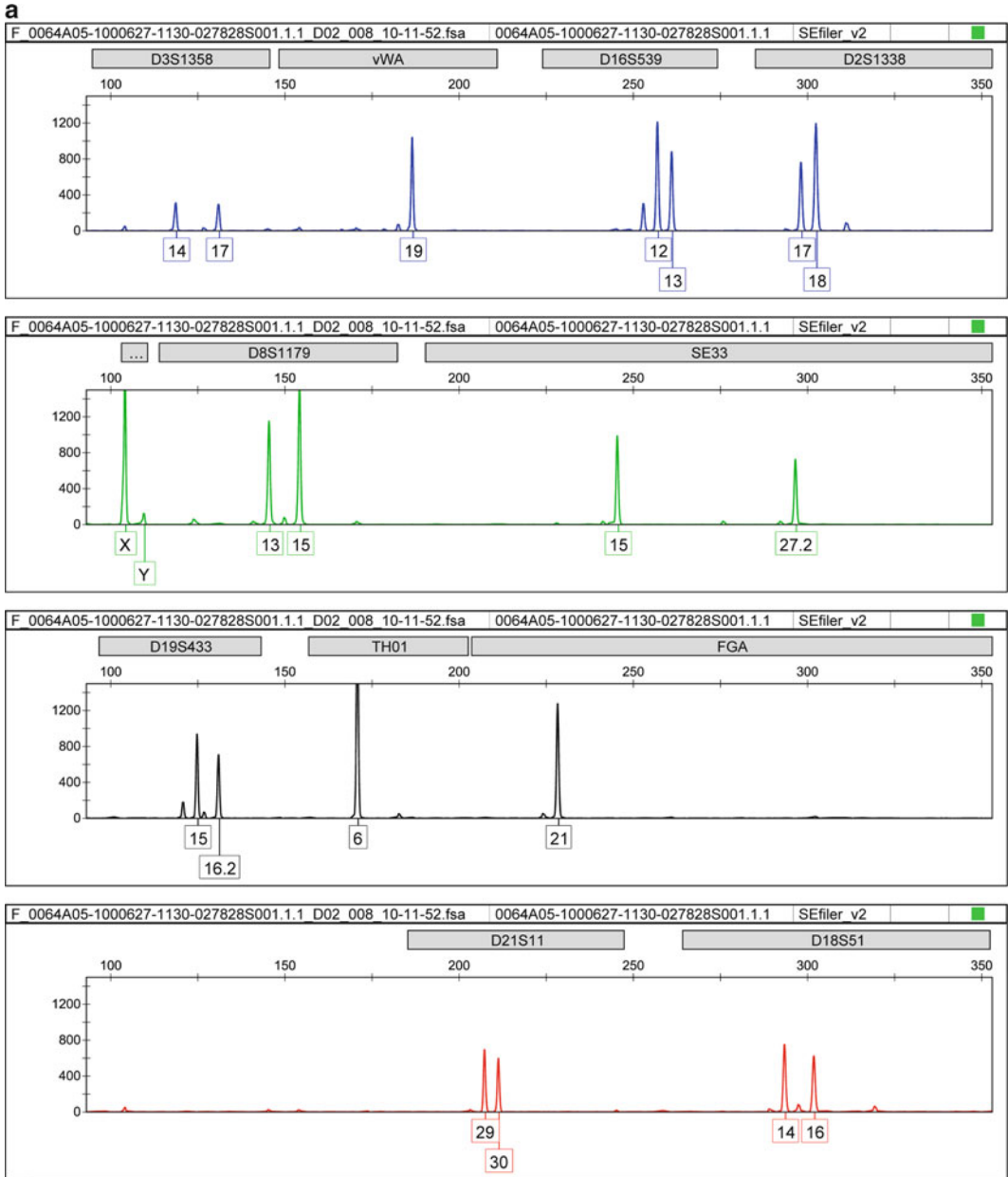
Fig. 1. (**a**) Standard 11-locus autosomal STR analysis of a DNA mixture. A predominant female DNA component is detected. (**b**) 17-locus Y-STR analysis of the same trace found at a condom. Despite the high female background, only the male DNA component is detected.

b
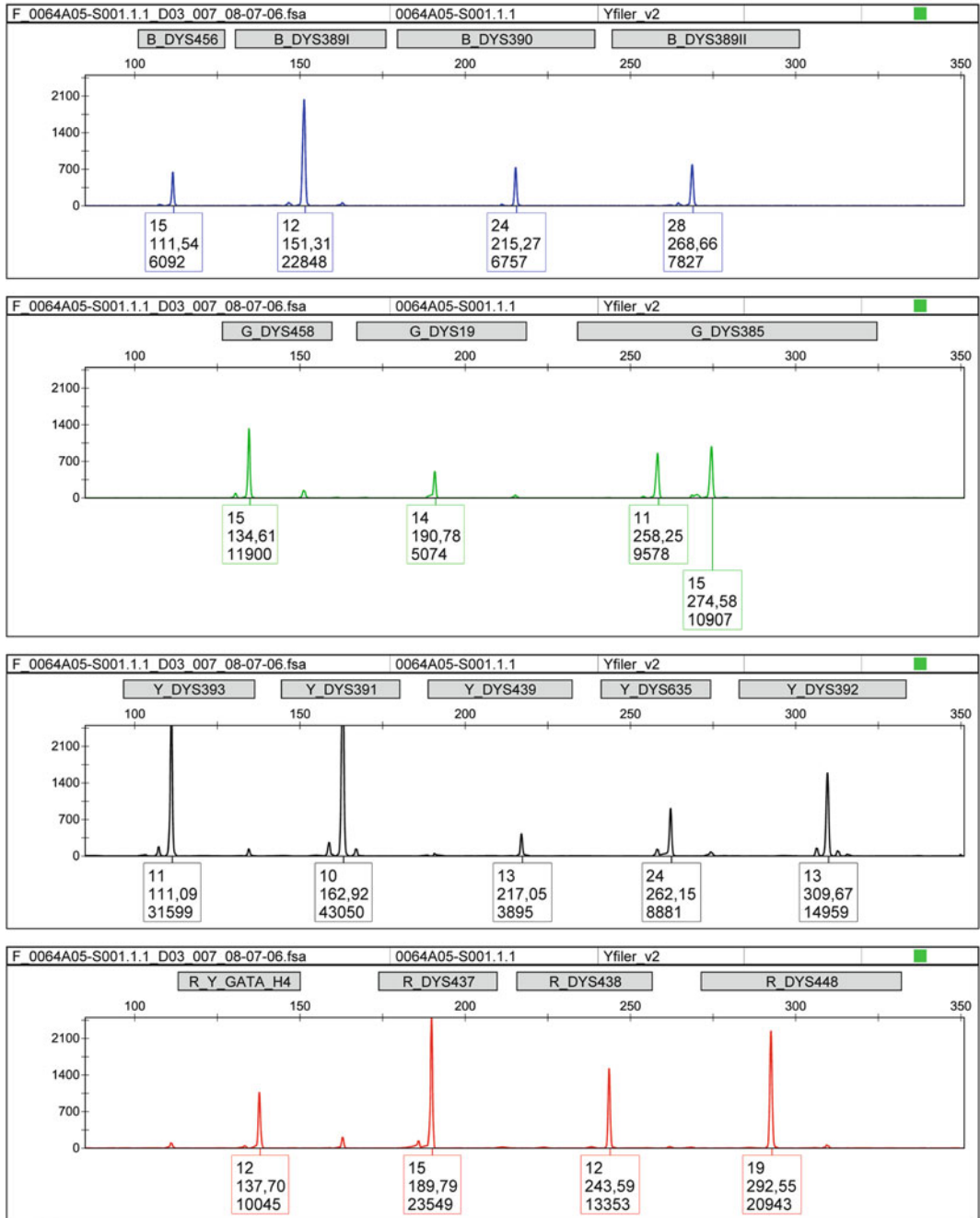
Fig. 1. (continued)

Table 1
Single-source 17-locus Y-STR haplotype

DYS 456	DYS 389I	DYS 390	DYS 389II	DYS 458	DYS 19	DYS 385ab	DYS 393	DYS 391	DYS 439	DYS 635	DYS 392	GATA H4	DYS 437	DYS 438	DYS 448
15	12	24	28	15	14	11,15	11	10	13	24	13	12	15	12	19

- Intermediate allele values such as "17.2" or duplicated (triplicated, quadruplicated) loci such as "14,15" (see the list of intermediate and multiple alleles for each locus at http://www.yhrd.org/Research/Loci).
- Null allele: enter "0" as the allele designation when the sample is believed to contain a null allele type created by mutation (and not by drop-out).
- Leave a particular locus blank (default) to find matches for any allele at this locus.
- To search at duplicated locus DYS385, enter the desired alleles separated by a comma, select the smallest allele first such as "11,14."
- Search in the "whole database" or select the ancestral group (metapopulation), a continent, or a national database of interest from the appropriate boxes. A list of the predefined metapopulations within the database is provided here http://www.yhrd.org/Research/Metapopulations (see Note 3).

2. Select the "search" button to query the database for the chosen haplotype.

Frequency and Confidence Interval Calculations

1. Haplotype frequencies can be estimated simply by counting or by extrapolation. Both methods rely on databases which comprise haplotypes sampled in populations. The latter approach also known as the "frequency surveying method" (4–6) is based upon the similarity and frequency relationships observed among haplotypes already included in the database. This approach ensures that rare haplotypes retain their high evidential power even when the database used for estimation is of realistic and therefore moderate size.

2. When using the counting method for the frequency estimation, a 95% confidence interval (CI) is added to the frequency value to correct for database size and sampling errors. Given the error percentile α (0.05 in the case of 95% CI), the number of observations n, and the database size N, we calculate the lower and upper CI limits using the exact Clopper–Pearson method (12):

$$\left[B^{-1}\left(\frac{\alpha}{2}, n, N-n+1\right), B^{-1}\left(\frac{1-\alpha}{2}, n+1, N-n\right) \right].$$

3. Since the counting method may reduce the evidentiary power inadequately in cases of nonobservation (estimates are limited by the database size and impaired by sampling effects), we recommend the extrapolation method where the haplotype in question does not enter into the estimation process. For the frequency of a haplotype in question, the so-called "surveying method" yields a beta-type posterior distribution with mean

(mean likelihood estimate) and mode points (maximum likelihood estimate). This posterior distribution is obtained by extrapolation from the structure and frequency of all other haplotypes in the relevant ancestral group (metapopulation). To extrapolate, search the haplotype first and click on the button "Frequency surveying estimates." The estimation process can only be conducted if at least full minimal haplotypes (9 loci) are entered and "whole database" is chosen. Estimates are currently available for 11 metapopulations, which reasonably meet the quality criteria (size, representativeness) (see Note 5). Two values are calculated, the "mean" and the "mode" of the a-posteriori haplotype frequency distribution. We recommend using the "mean" value as a robust estimator of the haplotype frequency. Please note that two estimates are calculated, one with the haplotype included in the databases, the other without inclusion of the haplotype. We recommend the latter to avoid an unnecessary reduction of the evidentiary power.

Query for the example haplotype (Fig. 1b, Table 1) with three different frequency estimations

1. The search result for the example haplotype is shown in Fig. 2a. The report (default) includes the number of matches in the whole database (0/27.532 17-locus haplotypes) and in 33 metapopulations, e.g., 0/9,470 Eurasians or 0/10,632 East Asians (YHRD release 34 built at November 19, 2010). Other groupings of the databased samples are also available, namely seven continental and 104 national databases, which can be selected according to needs of the users from the toolbar. The frequency of the haplotype in question can be reported in the most conservative way as the upper confidence interval limit (1.34×10^{-4}/total database) or by adding the unobserved haplotype to the database ($1/database + 1$) resulting in a more truthful frequency estimate (3.6×10^{-5}/total database). The latter approach also takes a very conservative position assuming that the next sampled haplotype for the database is the example haplotype.

2. In this example case, we recommend the extrapolation method because the origin of the DNA donor is known (Western Europe) and a large representative database is available for this ancestral pool. The mean frequency value is calculated with the program "Frequency Surveying Estimates" at the YHRD website. The value for the example haplotype is 7.0×10^{-5} in the Western European metapopulation (haplotype not included in the database) (Fig. 2b). Because the method is based on the structural relatedness of haplotypes within an ancestral pool, information on the origin of the DNA donor must be available a-priori. The extrapolation can only be done for metapopulations for which a large and representative spectrum of population-specific haplotypes is available in the YHRD. This was tested and confirmed
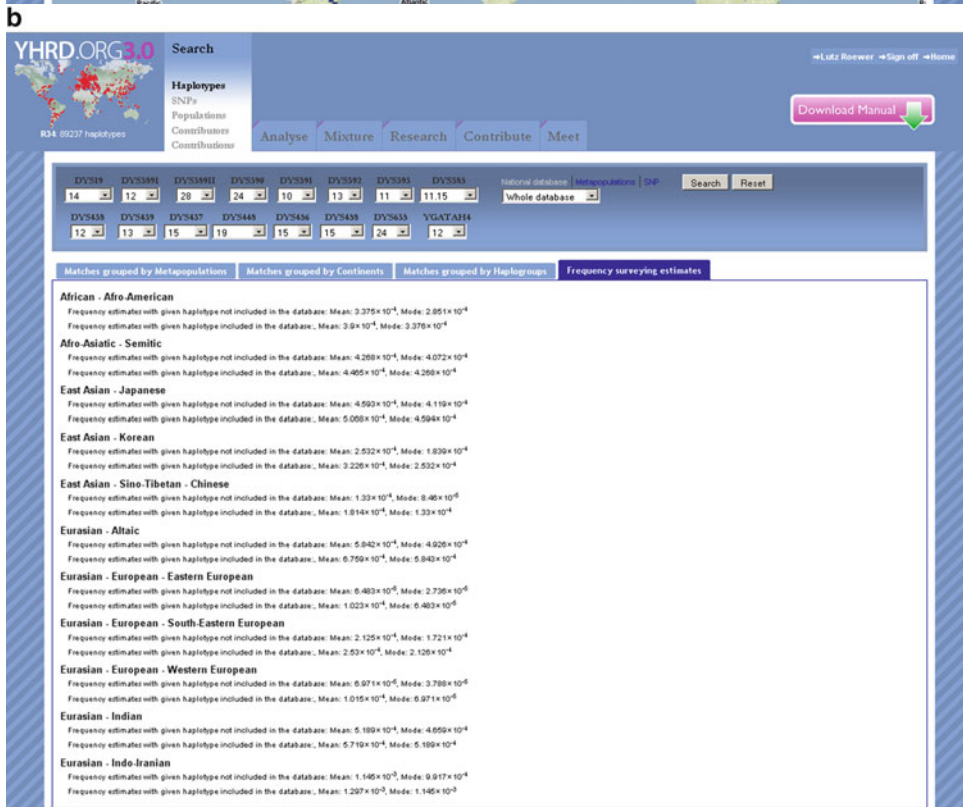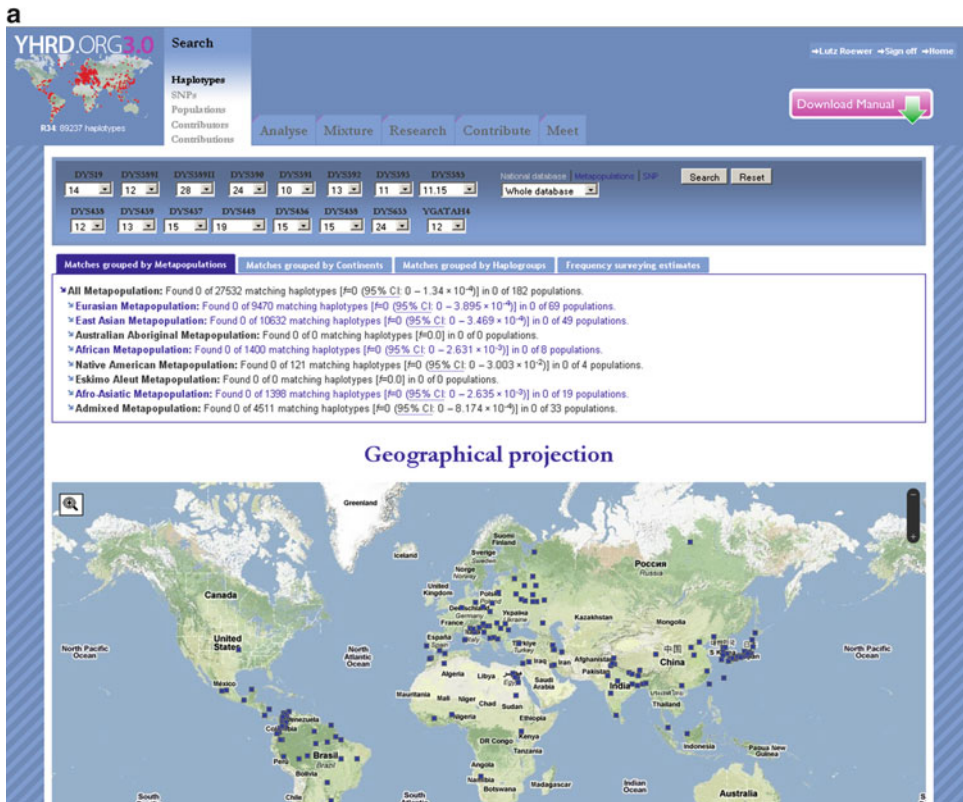
Fig. 2. (**a**) Frequency calculation (counting method) using the YHRD release 34 for the 17-locus profile amplified from the evidence. Matches are grouped according to metapopulations. (**b**) Frequency calculation using the frequency surveying method of the YHRD. Estimates are listed for different metapopulations.

for the Western European, the Eastern European, the Southeastern European, the Afro-American, the Japanese, the Korean, the Chinese, the Indian, Arab, Turkic, and the Indo-Iranian MP (6).

3.2. Male/Male Mixture Analysis

A case with a mixture consisting of more than one 17-locus Y-STR profile detectable in a trace is shown in Table 2, Fig. 3.

For the interpretation of such evidence, the YHRD offers a "mixture tool" (www.yhrd.org/Mixture) to calculate likelihood ratios for varying numbers of known and unknown male contributors (10). The database provides the population-dependent frequencies of the haplotype profiles detected in the trace.

1. Enter all alleles detected in the *trace* mixture separated by semicolons in the appropriate boxes (leave missing loci blank). Please note that all alleles at DYS385 have to be separated by semicolons as well.

2. Please enter the profile of the *putative donor* (e.g., suspect).

3. Choose the appropriate *metapopulation* to which all contributors belong.

4. Enter the number of *additional contributors*. This is the number of all hypothetical contributors to the trace minus one (the putative donor).

5. Press *calculate*.

The result will be a likelihood of donorship vs. nondonorship of the putative donor to the trace. In the example case, the likelihood ratio (based on the Eurasian metapopulation with 9,470 17-locus haplotypes, YHRD release 34) is 2.4×10^3 (see Fig. 4). So, it is highly probable that the man who left the major DNA component has contributed to the trace. Note the following limitation: the haplotype of the suspect must occur in the selected metapopulation. Otherwise the calculation is not possible. Also the calculation cannot be accomplished, if the selected database is too small. In this case, a failure message will appear (see Note 6).

4. Notes

1. A much more detailed description of the YHRD and its tools is published in the manual "YHRD 3.0, Directions for use," which can be downloaded from the website www.yhrd.org

2. Because the YHRD is updated regularly, sample sizes and frequencies change. The release number and date should thus be included in the query report.

3. A similar database like YHRD (and partly overlapping with it) exists in the USA with national population samples typed for 11–17 Y-STR loci (http://usystrdatabase.org).

Table 2
A mixture consisting of more than one 17-locus Y-STR profile detectable in a trace

DYS 456	DYS 389I	DYS 390	DYS 389II	DYS 458	DYS 19	DYS 385ab	DYS 393	DYS 391	DYS 439	DYS 635	DYS 392	GATA H4	DYS 437	DYS 438	DYS 448
14;18	12;13	23;25	28;30	15	15	11;14;15	13	10	10;11	22;23	11	10;13	14;16	10;11	19;20
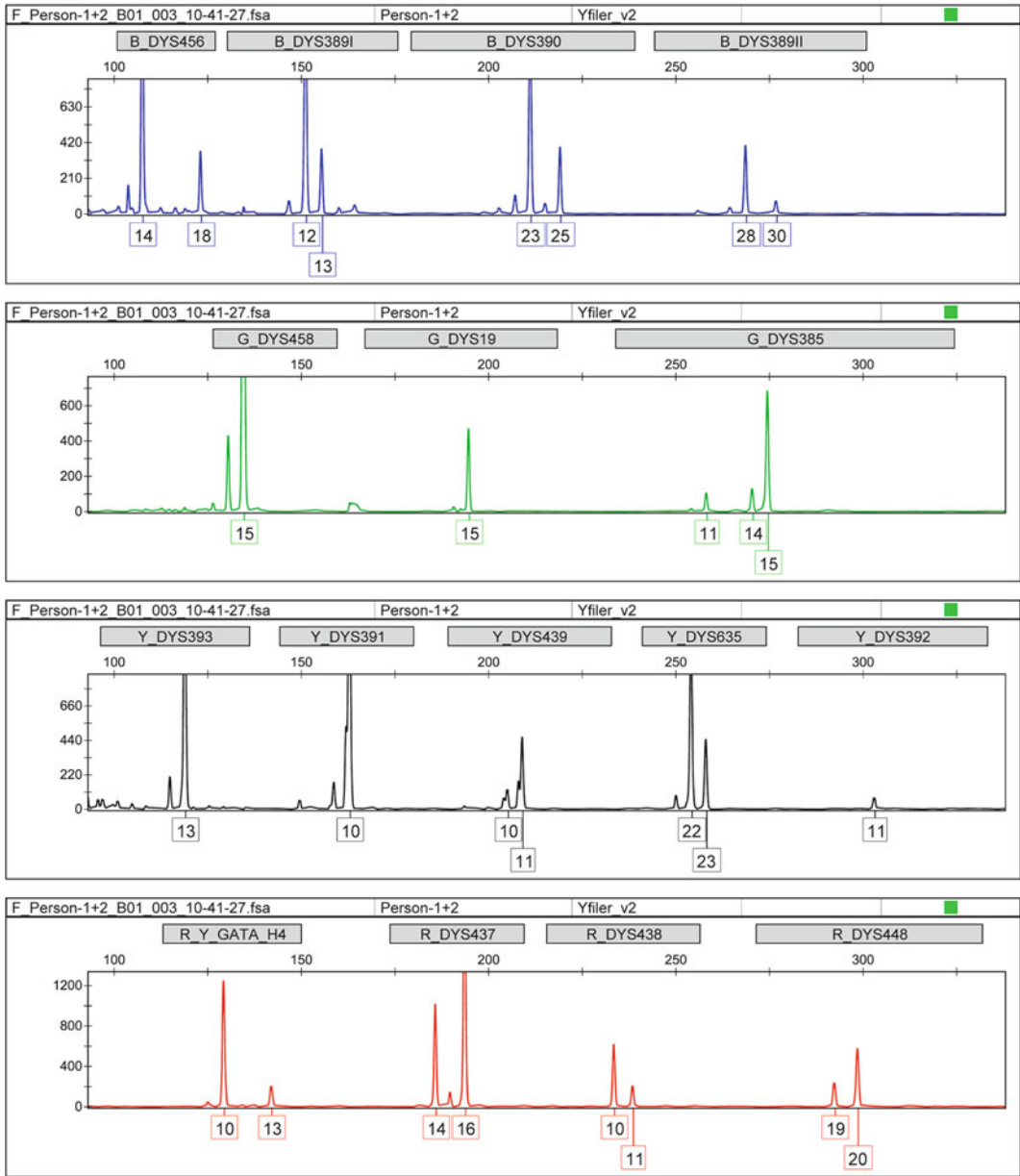
Fig. 3. 17-Locus Y-STR analysis of an un-balanced male/male DNA mixture.

4. While the different nature of underlying population structure has long been seen as an obstacle to combine the information obtained from lineage markers (Y and mtDNA) with recombining loci (autosomal STRs) into a single likelihood ratio, a recent seminal review re-examines this issue. According to the authors a combination of likelihoods is allowed when representative databases are used (13).

5. The public database YHRD relies on high scientific standards and transparency of the data collected within. All population

Fig. 4. Likelihood calculation using the online program "Mixture" of the YHRD.

studies submitted to the database are subject of rigorous quality assessment by the curators. All population studies which are accepted for publication in YHRD will be published in parallel in the most renowned forensic journals (8, 9). All forensic laboratories and institutions are invited to contribute to the YHRD. The resulting increase in the database size and the inclusion of geographically and ethnically varied population groups will augment the scientific and forensic efficacy of the database.

6. Various other applications (kinship, genealogical, DVI cases) can be solved with Y-STR analysis. The frequency estimation as outlined in this article is always the crucial step to interpret the results of such cases.

Acknowledgments

References

1. O'Donnell P, Montpetit S, Varlaro J, Schade L.L, Calandro L. (2008) Forensic Magazine, April/May, Vicon Publishing Inc., Amherst, USA.

2. Maiquilla SM, Salvador JM, Calacal GC, Sagum MS, Dalet MR, Delfin FC, Tabbada KA, Franco SA, Perdigon HB, Madrid BJ, Tan MP, De Ungria MC. (2011) Y-STR DNA analysis of 154 female child sexual assault cases in the Philippines. Int J Legal Med. 125(6):817–24.

3. Roewer L (2009) Y chromosome STR typing in crime casework. Forensic Sci Med Pathol 5: 77–84.

4. Roewer L, Kayser M, de Knijff P, Anslinger K, Betz A, Caglià A, Corach D, Füredi S, Henke L, Hidding M, Kärgel HJ, Lessig R, Nagy M, Pascali VL, Parson W, Rolf B, Schmitt C, Szibor R, Teifel-Greding J, Krawczak M (2000) A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. Forensic Sci Int. 114(1):31–43.

5. Krawczak M (2001) Forensic evaluation of Y-STR haplotype matches - a comment. Forensic Sci Int. 118(2–3):114–5.

6. Willuweit S, Caliebe A, Andersen MM, Roewer L (2011) Y-STR frequency surveying method: a critical reappraisal. For Sci International: Genetics 5(2):84–90.

7. Willuweit S, Roewer L; International Forensic Y Chromosome User Group (2007) Y chromosome haplotype reference database (YHRD): Update. Forensic Sci Int Genet 1(2):83–87.

8. Parson W, Roewer L (2010) Publication of population data of linearly inherited DNA markers in the International Journal of Legal Medicine. Int J Legal Med. 124(5): 505–509.

9. Carracedo A, Butler JM, Gusmão L, Parson W, Roewer L, Schneider PM. Publication of population data for forensic purposes (2010) Forensic Sci Int Genet. 4(3):145–7.

10. Wolf A, Caliebe A, Junge O, Krawczak M (2005) Forensic interpretation of Y-chromosomal DNA mixtures. Forensic Sci Int. 152(2–3):209–213.

11. M. Kayser, A. Cagliá, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, F. Heidorn, S. Herrmann, B. Herzog, M. Hidding, K. Honda, M. Jobling, M. Krawczak, K. Leim, S. Meuser, E. Meyer, W. Oesterreich, A. Pandya, W. Parson, G. Penacino, A. Perez-Lezaun, A. Piccinini, M. Prinz, C. Schmitt, P.M. Schneider, R. Szibor, J. Teifel-Greding, G. Weichhold, P. de Knijff, L. Roewer (1997) Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med. 110:125–133.

12. Clopper, C.J. and E.S. Pearson, The use of confidence or fiducial intervals illustrated in the case of the binomial. Biometrika (1934). 26: p. 404–413.

13. Buckleton JS, Krawczak M, Weir BS (2011) The interpretation of lineage markers. For Sci International: Genetics 5(2):78–83.

Chapter 5

Capillary Electrophoresis of an X-Chromosome STR Decaplex for Kinship Deficiency Cases

Leonor Gusmão, Cíntia Alves, Iva Gomes, and Paula Sánchez-Diz

Summary

During the two last decades, STR markers located on the autosomes have been gaining relevance and have nearly replaced the use of other type of markers in most cases of genetic identification, paternity testing, as well as in other situations of kinship analysis. Nevertheless, in some complex cases, independently of the number of polymorphisms being typed, autosomal markers convey very little information. Depending on the parentage constellation available for analysis, as well as the gender of the subjects, this problem can sometimes be solved by using markers that have different modes of transmission. Therefore, most forensic laboratories are nowadays prepared to analyse lineage markers (Y chromosome and mtDNA) and many have recently set up methods for the analysis of X-STRs. In the present chapter, a method is described for the typing of ten X chromosome-specific markers in a single PCR amplification reaction, followed by capillary electrophoresis separation and fluorescent detection in an ABI Genetic Analyser apparatus. This typing strategy was developed and optimized for the simultaneous amplification of ten X-linked specific STRs well distributed along the chromosome: DXS8378, DXS9902, DXS7132, DXS9898, DXS6809, DXS6789, DXS7133, GATA172D05, GATA31E08 and DXS7423.

Key words: X chromosome, X-STR, PCR multiplex, Paternity cases, Kinship analysis

1. Introduction

Due to their particular mode of transmission, genetic markers located on the non-recombining X-Y chromosome regions can represent an important source of information in some special cases of genetic identification or kinship analyses (1, 2), whenever the information from the autosomes is insufficient for an unambiguous evaluation of the alternative hypotheses in question.

Since the X chromosome-specific portion is present in a single – non-recombining – copy in males, it is transmitted from the father

to all daughters unchanged (unless a mutation occurs). Therefore, markers located on the X chromosome can be particularly useful in complex kinship cases where just few and/or distant related individuals are available for the genetic analysis (1, 2).

In the last few years, more than 50 X chromosome-specific STR (*Short Tandem Repeat*) polymorphisms have been described and the usefulness for parentage testing evaluated, namely in what concerns, sequence structure variation (3–6), mutation rates (7–9), allele frequencies in different populations and corresponding discrimination power and exclusion capacity (3, 7–11).

1.1. Genetic Transmission of the X Chromosome

As a result of the evolutionary processes that gave rise to the differentiation of the sex chromosomes in mammals (12), recombination between the X and Y chromosomes in humans is limited to two small portions in both tips of the short and long arms (the pseudoautosomal PAR1 and PAR2 regions) (see Fig. 1). The X chromosome transmission will depend on the gender of both the progenitor and the offspring (see Fig. 2). The father will pass on the single X chromosome copy only to daughters, without recombination (along most of the chromosome), and therefore all sisters (even from different mothers) must share at least one allele per locus for the X chromosome (unless mutation occurs). From the maternal side, after recombination, one of the two X chromosomes from the mother will be transmitted with fifty percent probability to both the male and female offspring, similar to autosomal marker transmission. Therefore, at each X chromosome locus, any two brothers will have the same probability of sharing or not the same maternal allele. This specific mode of transmission is responsible for some of the main advantages of the application of X chromosome loci in kinship cases, allowing an efficient complementation to the information provided by other type of genetic markers (1, 2).

1.2. The X chromosome in Kinship Analysis

In paternity cases (duos or trios) with female offspring, the Mean Exclusion Chance (MEC) for X chromosome STRs is higher than the MEC obtained for autosomal loci (1, 13) which considerably improves the level of efficiency by increasing the statistical power. However, in these scenarios, there is rarely the need to employ X chromosome markers since autosomal polymorphisms have the power to solve the case.

The main advantage of X chromosome STRs is based on their potential application in specific kinship cases normally when the analysis of autosomal markers is limited or insufficient (1, 2, 14–17), namely:

–   In a paternity case (duo or trio) when the child under dispute is of female gender and the set of autosomal STRs routinely used is not sufficiently informative. For example, when two or
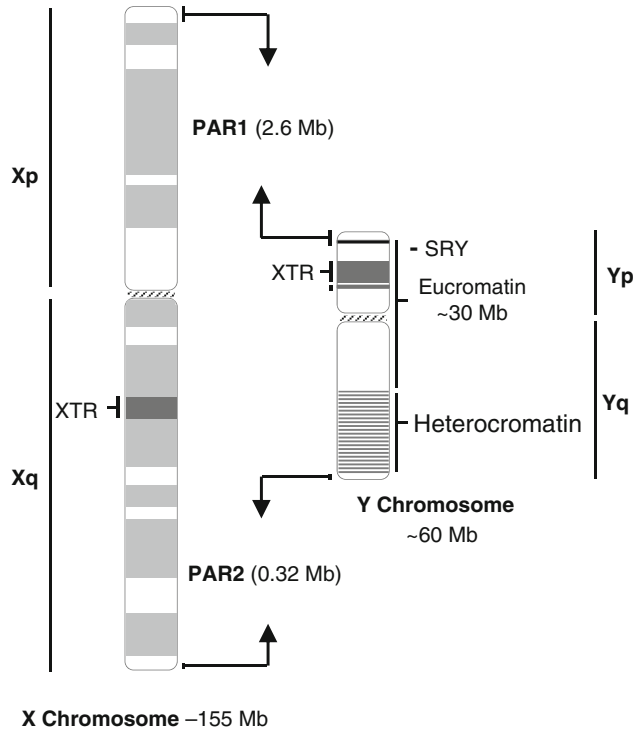
Fig. 1. Schematic representation of X and Y human chromosomes. *PAR1 and PAR2* pseudo-autosomal regions 1 and 2, *SRY* sex-determining region of the Y, *XTR* X-transposed region).
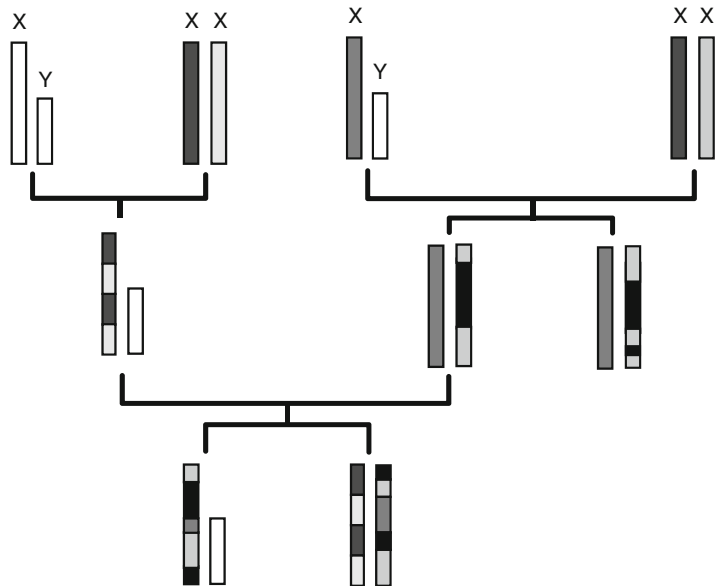
Fig. 2. X-chromosomal inheritance. Daughters inherit one paternal unchanged X chromosome because it does not undergo recombination (with exception for the pseudoauto-somal regions). After recombination, both daughters and sons inherit one of the maternal X chromosomes.

three father/daughter incompatibilities are observed in a total of 17 autosomal STRs analysed, the result may be considered inconclusive given that these occurrences may have one of two possible explanations: a real mutation or the true father is a close relative of the analysed individual. If additional autosomal STRs are not available, the analysis of X-chromosomal STRs can be used to increase the informativity and solve this specific case scenario.

– In paternity testing with a female descendant when the presumed fathers are close relatives, for example, father and son. In such cases, the analysis of X chromosome markers is highly informative, since father and son do not share identical alleles by descent (the son inherits the X chromosome from his mother and consequently father and son do not share any X chromosome allele).

– In paternity testing in sexual assaults with pregnancy interruption, the foetal material may be subjected to testing. In most of these cases, foetal material is contaminated by maternal component. If the foetus is of female gender, X chromosome markers can be used, in addition to autosomal markers, presenting higher power of exclusion a priori, increasing the informative power of the analyses.

– In maternity mother/son testing, X chromosome polymorphisms also present a higher MEC than autosomal STRs; this is the same for father/daughter paternity analysis. Although these types of scenarios are not so common, they are highly represented in mass disaster cases because identification of remains is usually achieved by testing the family members of the victim(s). In such cases, relationships such as mother/son (or mother/daughter) are considered to be more reliable since, in the present society, there is a high number of illegitimate paternities (registered fathers that are not the biological ones).

– Finally, in incomplete paternity testing, X-STRs are particularly valuable. Normally, these types of kinship testing refer to scenarios when the presumed father is not available for genetic analysis and, consequently, testing close blood relatives is necessary. In fact, such cases are very common because financial inheritance disputes are frequently the motif for an individual's request to prove his/her affiliation to the deceased alleged father or simply due to physical similarities (appearances) that substantiate the need to answer questions about possible common ancestry. The reencounter of family members is more significant in certain contexts such as wars, mass disasters and generalized international migration. In these cases, because paternal and/or maternal relatives must be studied, X chromosome markers are of great relevance due to their possible higher efficiency over autosomal markers (depending on the case). For example, if two

sisters or half sisters question their paternity, the analysis of X chromosome markers alone allows the exclusion of paternity without the need to use other family relatives. This result cannot be achieved when using autosomal markers. Every sister shares the paternal X chromosome haplotype and therefore the non-shared alleles are from maternal origin.

It is important to underline that in cases of incomplete paternity, the central key figure is the mother of the presumed and non-available father (presumed paternal grandmother). In fact, when the alleged paternal grandmother is tested, the kinship testing will not be an "incomplete" paternity analysis given that all the X chromosome alleles of the alleged father are present in his mother. To a certain extent, it is also possible to reconstruct the alleged paternal grandmother's profile through other sons and daughters. In this case, if the brothers of the alleged father have a different allele at a given locus, then this means that the mother's genotype is heterozygous at that locus. On the other hand, if they present identical alleles, the mother can either be homozygous or heterozygous for the given locus. In this case, if there are linked loci, the probability of homozygosity of the original locus will be assessed by haplotyping.

1.3. X Chromosome STR Markers and Typing Strategies

The use of the X chromosome markers only recently was introduced in some laboratories for routine parentage testing. So far, only a small number of STRs have been validated for forensic use when compared with the number of those currently available for the autosomes and the Y chromosome (1, 2). At present, 55 X-STRs have been included in the online available database, at www.chrx-str.org (18). However, there is only complete and extensive information available for 15 X chromosome polymorphisms, namely DXS6789, DXS6809, DXS7132, DXS7133, DXS7423, DXS8378, DXS9902, DXS9898, DXS10074, DXS10101, DXS10134, DXS10135, GATA172D05, GATA31E08 and HPRTB, mostly due to the fact that these markers are included in multiplex systems widely used by forensic laboratories. The DXS7132, DXS7423, DXS8378, DXS10074, DXS10101, DXS10134, DXS10135 and HPRTB loci are included in one of the first forensic genotyping kits for X chromosome STRs, the Argus X-8 Mentype® (Biotype AG) (19). This set of eight X-STRs is grouped into four pairs of closely located loci with low recombination rate among them. Each of these pairs is located in one of the four "core" sets of clusters, four distinct linkage groups located along the X chromosome (1, 19) (see Fig. 3): group 1 (Xp22)-DXS8378 and -DXS10135; group 2 (Xq11)-DXS7132 and -DXS10074; group 3 (Xq26)-HPRTB and -DXS10101; group 4 (Xq28)-DXS10134 and -DXS7423. Due to their genetic proximity, the markers belonging to the same linkage group must be treated as haplotypes and not as independent loci, as a matter of

X-DECAPLEX   **Mentype® Argus X-12**
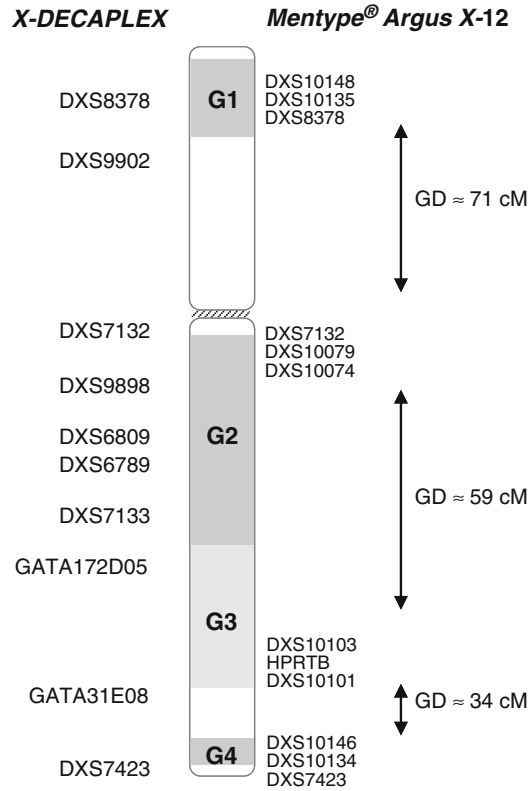
Fig. 3. Schematic representation of an X chromosome with location of the several commonly used X-STRs. On the left side of the chromosome are the STRs included in the X-Decaplex developed by the GHEP-ISFG (11). On the right are the markers included in the Investigator Argus X-12 (Qiagen, Hilden, Germany), with indication of the four linkage groups (G1–G4) (1, 19) and respective genetic distances (GD) between groups. GDs were determined using the online regression-based smoothed Rutgers map v.2 of the human genome (http://compgen.rutgers.edu/map-interpolator/) and using physical distances from NC_000023 human X chromosome reference sequence (NCBI build 36, last updated: 03 March 2008).

caution, when estimating the allele distribution in different populations and when calculating the powers of discrimination and exclusion a priori in paternity testing. More recently, the number of loci in this multiplex was increased to twelve, the Investigator Argus X-12 (Qiagen, Hilden, Germany) (see Fig. 3), by including an extra marker inside each of the four different linkage groups, namely DXS10148, DXS10079, DXS10103 and DXS10146.

The X chromosome markers DXS8378, DXS9902, DXS7132, DXS9898, DXS6809, DXS6789, DXS7133, GATA172D05, GATA31E08 and DXS7423 have also been widely studied for forensic applications (4, 5, 7, 10, 11). These loci are included in a multiplex system developed and optimized in a collaborative study of the GHEP-ISFG group (Spanish and Portuguese Speaking working group of the International Society for Forensic Genetics) (11).

In contrast to the genetic markers included in the above-mentioned commercial kits, the markers of the X-Decaplex are spread throughout the X chromosome at relatively far distances (see Fig. 3). This chapter will continue by focusing on the typing techniques and protocols necessary to analyse samples with this X Chromosome STR decaplex.

2. Materials

1. PCR reagents: QIAGEN Multiplex PCR kit (Qiagen); 10×
   Primer set (including labelled and non-labelled oligonucleotide
   primers listed in Table 1). The following disposable material is
   also required: plates and/or tubes and caps, gloves, pipette tips.
   All amplification reagents must be stored at –15 to –25°C in a
   constant temperature freezer. The labelled primer solutions are
   light-sensitive and therefore must be kept in the dark.

Table 1
Primer sequences and fluorescent dye labels

Locus	Primer sequence (5′–3′) and dye label	Ref.
DXS8378	6-FAM_TTAGGCAACCCGGTGGTCC ACAAGAACGAAACTCCAACTC	(3)
DXS9898	6-FAM_CGAGCACACCTACAAAAGCTG TAGGCTCACCTCACTGAGCA	(10)
DXS7133	6-FAM_CACTTCCAAAAGGGGAAAAA ACTTGTACTTGGTGGGAGGAA	(11)
GATA31E08	6-FAM_GCAAGGGGAGAAGGCTAGAA TCAGCTGACAGAGCACAGAGA	(11)
GATA172D05	VIC_TAGTGGTGATGGTTGCACAG ATAATTGAAAGCCCGGATTC	(3)
DXS7423	VIC_GTCTTCCTGTCATCTCCCAAC TAGCTTAGCGCCTGGCACATA	(3)
DXS6809	VIC_TCCATCTTTCTCTGAACCTTCC TGCTTTAGGCTGATGTGAGG	(10)
DXS7132	NED_TCCCCTCTCATCTATCTGACTG CACTCCTGGTGCCAAACTCT	(10)
DXS9902	NED_CTGGGTGAAGAGAAGCAGGA GGCAATACACATTCATATCAGGA	(11)
DXS6789	NED_CTTCATTATGTGCTGGGGTAAA ACCTCGTGATCATGTAAGTTGG	(10)

2. Analysis of the amplified products can be performed using any ABI Genetic Analyzer instrument (AB Applied Biosystems, Foster City, CA). The following material is described for an ABI 3130 Genetic Analyzer (according to the instrument manufacturer's instructions): plates and septa (AB Applied Biosystems); internal size standard GeneScan™-600 LIZ (AB Applied Biosystems); dry heating block or thermal cycler; capillaries; Performance Optimized Polymer 7, POP-7 (AB Applied Biosystems); Hi-Di formamide (AB Applied Biosystems); and ice.

Note that:

– To prevent cross-contaminations, it is recommended to use disposable gloves and aerosol-resistant pipette tips. Pre- and post-PCR samples, reagents, and instruments should be used and stored separately.

– If PCR amplification reagents are on the workbench for a long period while setting up the reaction, it is suggested to keep the solutions on ice, especially if room temperature is high.

– The fluorescent dyes attached to the primers and size standard are light-sensitive. Primer solutions, amplified DNA and GeneScan™-600 LIZ should be protected from light particularly when not in use.

– Formamide is potentially hazardous (irritant and a teratogen) and should be handled accordingly.

3. Methods

The following method described consists in the simultaneous PCR amplification of ten X chromosome STRs. This X-Decaplex method was optimized in a GeneAmp® PCR System 9700 Thermal Cycler (AB Applied Biosystems) and detection performed on an ABI PRISM®3130 Genetic Analyzer. Note that protocols may need to be optimized for different instruments and even the same instrument model may need slight protocol adjustments.

3.1. Loci Information   The loci included in the X-Decaplex were selected taking into account the gene diversity values reported in different populations (e.g. (4, 20–28)) and the potential for multiplexing. Preference was given to simple rather than complex STRs, following the ISFG recommendations concerning locus selection for forensic applications (29). In order to reduce the potential PCR-generated slippage artefacts, loci comprising trinucleotide repeats were avoided. In Table 2, the repeat sequence structures and the allele range for the loci included in the X-STR decaplex here described are showed.

Table 2
Repeat structure of the ten X-STRs included in the Decaplex, and corresponding references

Locus	STR sequence structure	Ref.	Allele range	Size range (bp)
DXS8378	$(CTAT)_n$	(20)	8–14	106–134
DXS9902	$(TAGA)_n$	(11)	8–16	170–202
DXS7132	$(CTAT)_n (CAT)_{0-1} (CTAT)_{0-2}$	(20)	9–18.3	116–155
DXS9898	$(TATC)_2$ atc $(TATC)_n (ATC)_{0-1} (TATC)_m$	(23)	7–16	145–181
DXS6809	$(CTAT)_n (ATCT)_3$ atcatctat $(TATC)_m$ $(ATCT)$ attatctatc $(ATCT)_p$	(21)	27–39.1	230–279
DXS6789	$(TATC)_{0-1} (TATG)_n (TATC)_m$	(22)	10–26	233–297
DXS7133	$(ATAG)_n$	(20)	6–14	181–213
GATA172D05	$(TAGA)_n$	(3)	6–17	209–253
GATA31E08	$(AGGG)_{2-3} (AGAT)_n$	(4)	7–16	224–260
DXS7423	$(TCCA)_3$ (tctgtcct)$_{0-1}$ $(TCCA)_n$	(28)	8–18	159–199

In small caps are indicated the bases not included in the repeat number counting for allele nomenclature purposes

3.2. PCR Amplification   Different technical methods were investigated with the aim of optimizing the multiplex systems for the simultaneous PCR amplification and genotyping of the following ten X chromosome STRs: DXS8378, DXS9902, DXS7132, DXS9898, DXS6809, DXS6789, DXS7133, GATA172D05, GATA31E08 and DXS7423. The best results were obtained when using between 0.5 and 5 ng template DNA in a 10 µL reaction volume containing 1× Qiagen multiplex PCR master mix, and all primers at a concentration of 0.2 µM.

3.2.1. Primer Mix Setup   Primer sequences and references are listed in Table 1. For those systems with overlapping allele size ranges, the primers were labelled with different dyes (see allele size ranges in Table 2) to allow an unambiguous typing of all the markers (see Fig. 4). Various primer concentrations were tested and the best co-amplification results were obtained when using all primers at a concentration of 0.2 µM. A 10× primer mix can be prepared for a large number of reactions and stored at –15 to –20°C until needed for PCR setup. From initial 100 µM primer stocks, an X-Decaplex 10× primer mix can be prepared, for example for 100 PCR amplification reactions by pipetting 2 µL of each forward and reverse primers and by adding 60 µL of sterile molecular biology grade water for completing 100 µL final volume. For studies requiring a large number of samples to be typed (e.g. population studies), it is recommended to
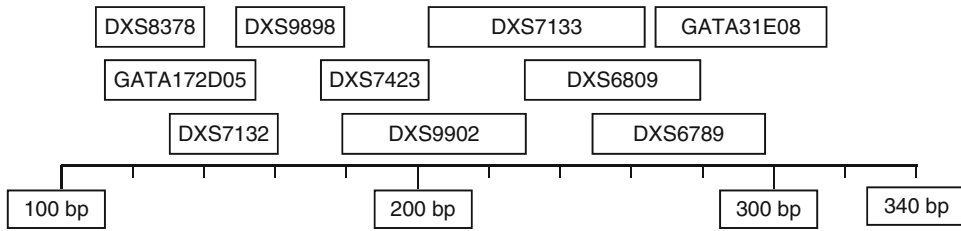
Fig. 4. Schematic representation of the size range distribution of the X-STRs included in the X-Decaplex. *First line*: markers labelled with 6-FAM; *second line*, with VIC and; *third line*, with NED (see Table 1 for primer details, and Table 2 for alleles' size ranges).

prepare larger volumes of 10× primer mix in a single tube, mix the final solution by thoroughly vortexing, and distribute smaller volumes into different tubes, before storing at –15 to –20°C (e.g. prepare a primer mix of 500 μL and distribute 50 μL by ten tubes, each one containing enough volume for typing 50 samples). This has the advantage of reducing pipetting errors for small volumes and allows better homogeneity between different samples, as well as in different PCR reactions for the same DNA sample. It also reduces possible effects on primer quality by less freeze/defreeze events and less loss of product will occur if necessary to discard the primer mix solution, for example, due to contamination.

3.2.2. Amplification Procedure

To set up the amplification, prepare a PCR master mix for all samples by determining the number of reactions and adding one or two reactions to this number to compensate for pipetting errors.

According to the stock concentrations, the following volumes are indicated for the PCR reagents per sample reaction (1×):

2× Qiagen Multiplex PCR Master Mix	5 μL
10× Primer mix (2 μM of each primer)	1 μL
Distilled water	3.5 μL
Template DNA (1–10 ng/μL)	0.5 μL
	Final volume = 10 μL

If stock solutions with different concentrations are used, the volumes must be adjusted to keep the same final concentration of each reagent in the final reaction solution.

Note that the total final reaction volume should be 10 μL and that changes in template DNA volume should be compensated by changing the volume of water volume (see Notes 1 and 2). The inclusion of a "negative control" (PCR mix plus water instead of DNA template) is important for detection of possible PCR contamination by external DNA.

3.2.3. Thermal Cycling Conditions

The thermal cycling conditions were optimized in a GeneAmp® PCR System 9700 Thermal Cycler (AB Applied Biosystems). Although significant changes to PCR performance are not expected, the use of other instruments may require some adjustments on denaturing, annealing, and extension steps. The annealing temperature is the most frequent parameter varying between apparatus that may need some adjustment, because lowering annealing temperatures usually increases the amplification of unspecific products.

The number of cycles can also be adjusted for different input DNA (see Note 1). In forensic DNA samples, it is not recommended to increase the total number of cycles for more than 35–40 cycles.

Thermo cycling conditions are: pre-incubation for 15 min at 95°C; followed by 10 cycles of 30 s at 94°C, 90 s at 60°C, 60 s at 72°C; and 20 cycles of 30 s at 94°C, 90 s at 58°C, 60 s at 72°C; with a final incubation for 60 min at 72°C (see Note 3).

3.3. Analysis of Amplified Fragments

The detection analysis was optimized on an ABI PRISM® 3130 Genetic Analyser (AB Applied Biosystems). When using other ABI platforms, consult the instrument's user's manual for appropriate adjustments on sample preparation and electrophoresis conditions.

3.3.1. Sample Preparation

For sample preparation, it is recommended to make one loading mixture for all the samples to be analysed (13.75 µL of Hi-Di formamide and 0.25 µL of LIZ 600 internal size standard per sample). When preparing the loading mixture, add one or two samples extra in the calculations to compensate for pipetting errors. Dispense 14 µL of the loading mixture into each ABI 3130 plate slots and add 1 µL of the PCR product. Before loading, denature the samples by heating them at 95°C for 3 min and immediately placing them on ice for at least 5 min.

3.3.2. Analysis Conditions

Separation and detection can be performed in various types of ABI Genetic Analysers (Applied Biosystems), following manufacturer's instructions. On an ABI PRISM® 3130 Genetic Analyzer, samples are analysed by electrophoresis using Performance Optimized Polymer 7 (POP-7; AB Applied Biosystems) and filter set G5. Although this multiplex was optimized for typing with an internal size standard GeneScan-600 LIZ (Applied Biosystems) and filter G5, other platforms allowing for the detection of dyes with maximum fluorescence at only four different wavelengths may be used: size standard GeneScan-600 ROX and Filter D are recommended.

Fragment sizes are determined automatically using the GeneMapper ID® Analysis Software, version 3.2, and samples should be typed by comparison with sequenced allelic ladders. The use of sequenced ladders is especially important for DXS9898, DXS9902 and DXS7132 (see Notes 4 and 5) in order to allow the unequivocal typing of intermediate alleles detected in most populations.

Table 3
X-STR profiles of standard DNA samples used as references by the participating laboratories

	9947A	9948	NA3657	Refs.
DXS8378	10/11	11	12	(30)
DXS9898	12/15	13	–	(4)
DXS7133	9/10	11	9	(30)
GATA31E08[a]	13	12	–	(11)
GATA172D05	10	6	9	(30)
DXS7423	14/15	14	13	(30)
DXS6809	31/34	31	29	(30)
DXS7132	12	13	12	(30)
DXS9902[b]	12	13	13	(11)
DXS6789	21/22	20	23	(30)

[a]Genotypes were changed by adding two repeats to the previously reported by Shin et al. (31)
[b]Genotypes were changed by adding one repeat to the previously reported by Szibor et al. (30)

It is important to use DNA reference samples (e.g. 9947A, 9948 and NA3657) as genotyping controls, which have been previously typed for all markers included in the X-Decaplex (see Table 3).

3.4. Allelic Ladders The use of sequenced allelic ladders is very successful for comparison purposes in forensic and other studies, and it is thus recommended by the DNA Commission of the International Society for Forensic Genetics (32). The general strategy for ladder construction is primarily to isolate and sequence different-sized STR alleles found in a first screening of a population sample. After identifying the different alleles by sequence analysis, the samples can be mixed and amplified together to produce a ladder. The volume of each individual sample in the mixture must be adjusted to produce a well-balanced ladder. For multiplex analysis, the best strategy is to produce single ladders for each locus and mix them together to make up the final multiplex ladder. Single-locus ladders can be re-amplified using 1 μL (in a 50-μL final reaction volume) of between 1:1000 and 1:10.000 dilutions of the original ladder (see Note 6).

4. Notes

Besides those already mentioned, some other problems can arise during the X-Decaplex set up, as follows:

1. Peak heights between 1,000 and 2,000 relative fluorescent units are ideal. If peak heights are too high for some loci, reduce the PCR primer concentration of those loci. If this happens to all the loci included in the multiplex, use less DNA or reduce the number of PCR cycles.

2. An excess of DNA during PCR will improve the amplification of smaller-sized loci and a reduction in the amplification of larger Alleles can be observed.

3. An amplified fragment may show a second peak (–1 bp) because of the incomplete 3' adenine base addition. This artefact can be reduced with the increase of the PCR final extension step or by switching the dye label to the other primer.

4. For DXS9902 locus, intermediate alleles can be present due to a T insertion in a tract of 9 Ts ending 5 bases upstream the GATA repeat (5). To ensure the correct typing of these alleles, it is very important to use sequenced ladders including the most frequent intermediate alleles 12.1 and 13.1.

5. Allele 8.3 at DXS9898 is described as having a high frequency in all European populations studied until now (e.g. (23–25, 28, 33)). In a previous study by Gusmão et al. (11), it was observed that allele 8.3 differed in approximately 6 bp from allele 10, and allele 7 differed in approximately 4 bp from allele 8.3, which can lead to genotyping errors if sequenced allelic ladders are not used. The same kind of typing problem was observed for some DXS7132 alleles, in the same size range of DXS9898 alleles 7 and 8.3. This was found to be caused by a sizing problem when using GeneScan-500 LIZ size standard for fragments in the range of 139–150 bp. This is why we recommend the use of GeneScan-600 LIZ size standard which does not show this problem. Whichever size standard is chosen, to ensure the correct typing of these two STRs, sequenced ladders should be used, including the most frequent alleles at these loci.

6. Amplification of stutter products (with one repeat less) can appear because of the enzyme slippage during amplification. For the multiplex described here, no significant stutter peaks were observed. However, special attention should be given in successive re-amplification of the ladders. Usually, with ladder re-amplification, the larger alleles amplify less efficiently with the simultaneous increase of stutter amplification. This effect has been sometimes responsible for a one repeat less drift in

the correct allele typing. Therefore, after re-amplification, the ladder should be tested by comparison with the original one or by using sequenced control samples.

The X-specific STR typing by PCR amplification and fragment size determination follows the same principles as autosomal STR typing. Therefore, recommendations concerning STR typing should be followed, including samples, apparatus, and reagent handling (see other chapters on STR typing). User's manuals for commercial STR multiplex amplification kits can also be useful for consulting additional topics on troubleshooting.

Acknowledgments

References

1. Szibor, R., Krawczak, M., Hering, S., Edelmann, J., Kuhlisch, E., and Krause, D. (2003) Use of X-linked markers for forensic purposes. *Int J Legal Med* 117, 67–74.

2. Szibor, R. (2007) X-chromosomal markers: past, present and future. *Forensic Sci Int Genet* 1, 93–99.

3. Edelmann, J., Deichsel, D., Hering, S., Plate, I. and Szibor, R. (2002) Sequence variation and allele nomenclature for the X-linked STRs DXS9895, DXS8378, DXS7132, DXS6800, DXS7133, GATA172D05, DXS7423 and DXS8377. *Forensic Sci Int* 129, 99–103.

4. Gomes, I., Prinz, M., Pereira, R., Bieschke, E., Mayr, W.R., Amorim, A., Carracedo, A. and Gusmão, L. (2009) X-chromosome STR sequence variation, repeat structure, and nomenclature in humans and chimpanzees. *Int J Legal Med* 123, 143–149.

5. Gomes, I., Pereira, R., Mayr, W.R., Amorim, A., Carracedo, A. and Gusmão, L. (2009) Evaluation of DXS9902, DXS7132, DXS6809, DXS7133, and DXS7423 in humans and chimpanzees: sequence variation, repeat structure, and nomenclature. *Int J Legal Med* 123, 403–412.

6. Szibor, R., Edelmann, J., Hering, S., Gomes, I. and Gusmão, L. (2009) Nomenclature discrepancies in the HPRTB short tandem repeat. Int J Legal Med 123, 185–186.

7. Pico, A., Castillo, A., Vargas, C., Amorim, A. and Gusmão, L. (2008) Genetic profile characterization and segregation analysis of 10 X-STRs in a sample from Santander, Colombia. *Int J Legal Med* 122, 347–351.

8. Fracasso, T., Schurenkamp, M., Brinkmann, B. and Hohoff, C. (2008) An X-STR meiosis study in Kurds and Germans: allele frequencies and mutation rates. *Int J Legal Med* 122, 353–356.

9. Tillmar, A.O., Mostad, P., Egeland, T., Lindblom, B., Holmlund, G. and Montelius K. (2008) Analysis of linkage and linkage disequilibrium for eight X-STR markers. *Forensic Sci Int Genet* 3, 37–41.

10. Gomes, I., Prinz, M., Pereira, R., Meyers, C., Mikulasovich, R.S., Amorim, A., Carracedo, A. and Gusmão, L. (2007) Genetic analysis of three US population groups using an X-chromosomal STR decaplex. *Int J Legal Med* 12, 198–203.

11. Gusmão, L., Sánchez-Diz, P., Alves, C. et al (2009) A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations. *Int J Legal Med* 123, 227–234.

12. Lahn, B.T. and Page, D.C. (1999) Four evolutionary strata on the human X chromosome. *Science* 286, 964–967.

13. Krawczak, M. (2007) Kinship testing with X-chromosomal markers: Mathematical and statistical issues. *Forensic Science International: Genetics* 1, 111–114.

14. Edelmann, J., Lessig, R., Klintschar, M. and Szibor, R. (2004) Advantages of X-chromosomal

microsatellites in deficiency paternity testing: presentation of cases. *International Congress Series* 1261, 257–259.

15. Toscanini, U., Berardi, G., Gómez, A. and Raimondi, E. (2009) X-STRs analysis in paternity testing when the alleged father is related to the biological father. *Forensic Science International: Genetics Supplement Series* 2, 234–235.

16. Toni, C., Presciuttini, S., Spinetti, I., Rocchi, A. and Domenici, R. (2006) Usefulness of X-chromosome markers in resolving relationships: Report of a court case involving presumed half sisters. *International Congress Series* 1288, 301–303.

17. Silva, D.A., F.S.N, Manta, M., Desidério, C., Tavares and E.F., de Carvalho (2009). Paternity testing involving human remains identification and putative half sister: Usefulness of an X-hexaplex STR markers. *Forensic Science International: Genetics Supplement Series* 2, 230–231.

18. Szibor, R., Hering, S., Edelmann, J. (2006) A new Web site compiling forensic chromosome X research is now online. *Int J Legal Med* 120, 252–254.

19. Becker, D., Rodig, H., Augustin, C. J., Edelmann, F., Gotz, S., Hering, S., Szibor, R., and Brabetz W. (2008). "Population genetic evaluation of eight X-chromosomal short tandem repeat loci using Mentype Argus X-8 PCR amplification kit. *Forensic Sci Int Genet* 2, 69–74.

20. Edelmann J, Hering S, Michael M et al. (2001) 16 X-chromosome STR loci frequency data from a German population. *Forensic Sci Int* 124, 215–218.

21. Edelmann, J., Deichsel, D., Plate, I., Kayser, M., Szibor, R .(2003) Validation of the X-chromosomal STR DXS6809. *Int J Legal Med* 117, 241–244.

22. Hering, S., Kuhlisch, E., Szibor, R. (2001) Development of the X-linked tetrameric microsatellite marker HumDXS6789 for forensic purposes. *Forensic Sci Int* 119, 42–46.

23. Hering, S. Szibor R. (2000) Development of the X-linked tetrameric microsatellite marker DXS9898 for forensic purposes. *J Forensic Sci* 45, 929–931.

24. Pereira, R., Gomes, I., Amorim, A., Gusmão, L. (2007) Genetic diversity of 10 X-chromosome STRs in northern Portugal. *Int J Legal Med* 121, 192–197.

25. Poetsch, M., Sabule, A., Petersmann, H., Volksone, V., Lignitz, E. (2006) Population data of 10 X-chromosomal loci in Latvia. *Forensic Sci Int* 157, 206–209.

26. Shin, K.J., Kwon, B.K., Lee, S.S. et al. (2004) Five highly informative X-chromosomal STRs in Koreans. *Int J Legal Med* 118, 37–40.

27. Shin, S.H., Yu, J.S., Park, W.S, Min, G.S., Chung, K.W. (2005) Genetic analysis of 18 X-linked short tandem repeat markers in Korean population. *Forensic Sci Int* 147, 35–41.

28. Zarrabeitia, M.T, Amigo, T., Sanudo, C., Zarrabeitia, A., González-Lamuño, D., Riancho, J.A. (2002) A new pentaplex system to study short tandem repeat markers of forensic interest on X chromosome. *Forensic Sci Int* 129, 85–89.

29. Gusmão, L., Butler., J.M,. Carracedo, A. et al. (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Int J Legal Med* 120, 191–200.

30. Szibor, R., Edelmann, J., Hering, S. et al. (2003) Cell line DNA typing in forensic genetics: the necessity of reliable standards. *Forensic Sci Int* 138, 37–43.

31. Shin, S.H., Yu, J.S., Park, W.S., Min, G.S., Chung, K.W. (2005) Genetic analysis of 18 X-linked short tandem repeat markers in Korean population. *Forensic Sci Int* 147, 35–41.

32. Bär, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P., Lincoln, P., et al. (1997) DNA recommendations-further report of the DNA Commission of ISFG regarding the use of short tadem repeat systems. *Int J Legal Med* 11, 175–176.

33. Aler, M., Sánchez-Diz, P., Gomes, I., Gisbert, M., Carracedo, A., Amorim, A., Gusmão, L. (2007) Genetic data of 10 X-STRs in a Spanish population sample. *Forensic Sci Int* 173, 193–196.

Chapter 6

A 48-plex Autosomal SNP GenPlex™ Assay for Human Individualization and Relationship Testing

Carmen Tomas, Claus Børsting, and Niels Morling

Abstract

SNPs are being increasingly used by forensic laboratories. Different platforms have been developed for SNP typing. We describe the GenPlex™ HID system protocol, a new SNP-typing platform developed by Applied Biosystems where 48 of the 52 SNP*for*ID SNPs and amelogenin are included. The GenPlex™ HID system protocol has been successfully tested by a number of forensic laboratories using both ordinary and forensic samples.

Key words: Autosomal SNPs, SNP*for*ID consortium, GenPlex™, OLA reaction, ZipChute™ probes

1. Introduction

In recent years, the usefulness of SNPs in forensic genetics has been pointed out and an increasing number of laboratories are implementing SNP typing as a complementary analysis to STR typing (1–6). Some of the properties of SNPs make them especially suitable for forensic applications (a) SNPs can be analyzed from short amplicons, which makes it possible to analyze SNPs in partly degraded DNA samples (7–9) and (b) SNPs have a low mutation rate, which is an important advantage in kinship analysis (1). Many platforms can be used for the analysis of SNPs (3, 10–13). The most widely used one is the single base extension assay (SBE) using the SNaPshot® kit (Applied Biosystems). However, the SNaPshot® kit has some drawbacks that can make it a challenge to analyze the results (2). Applied Biosystems developed the GenPlex™ HID system, which is a new SNP-typing platform based on a modification of the SNPlex™ System (14). The assay includes 48 of the 52 SNPs in the SNP*for*ID 52-plex (15) and amelogenin. Briefly, the

GenPlex™ HID system protocol includes (see Fig. 1) an initial multiplex PCR reaction followed by a PCR reaction clean-up with Exonuclease I and Shrimp Alkaline Phosphatase. An oligo ligation assay (OLA) using a locus and two allele specific oligonucleotides (LSOs and ASOs) is performed on the purified PCR products. The biotinylated OLA products are bound to a solid phase, and ZipChute™ probes are hybridized to the zipcode sequences of the OLA products. After elution, the ZipChute™ probes are analyzed by capillary electrophoresis (CE). Each zipchute has a unique color/mobility characteristic and they can be uniquely identified by analysis of the CE data. There are two zipchutes per SNP, one for each of the two possible alleles. A number of laboratories have successfully tested the GenPlex™ HID System and demonstrated that it is a sensitive and reproducible SNP-typing method (16, 17). Also, full or nearly full profiles could be obtained when samples with partially degraded DNA were analyzed ((18), see Fig. 2).

2. Materials

Some of the reagents used in this protocol are hazardous and must be handled with appropriate clothing and gloves. Waste materials must be discarded in the appropriate waste containers. Follow always the security and storage recommendations of the manufaturer.

2.1. PCR Reaction

1. GenPlex™ Multiplex PCR Primer Mix (Applied Biosystems, Foster City, CA, USA). Store at –15 to –25°C.
2. QIAGEN Multiplex PCR Kit (catalog number 206143, Qiagen GmbH, Hilden, Germany). Store at –15 to –25°C.
3. Nuclease-free water.

2.2. PCR Product Clean-up

1. ExoSAP-IT® (catalog number 78201, USB Corp., Cleveland, OH, USA). Store at –15 to –25°C.

2.3. Phosphorylation of the Oligonucleotide Ligation Assay Oligos and Oligonucleotide Ligation Assay Reaction

1. Oligonucleotide Ligation Assay (OLA) Oligo Mix (catalog number 4383847, Applied Biosystems). Store at –15 to –25°C.
2. OLA Master Mix SNPlex™ System (catalog number 4362268, Applied Biosystems). Store at –15 to –25°C.
3. dATP (100×) SNPlex™ System (catalog number 4362268, Applied Biosystems). Store at –15 to –25°C.

2.4. Immobilization of the OLA Products on the Hybridization Plate

1. SNPlex™ System Hybridization Plates, 96-Well (catalog number 4357279, Applied Biosystems).
2. Hybridization Wash Buffer (10×) SNPlex™ System (catalog number 4349301, Applied Biosystems).
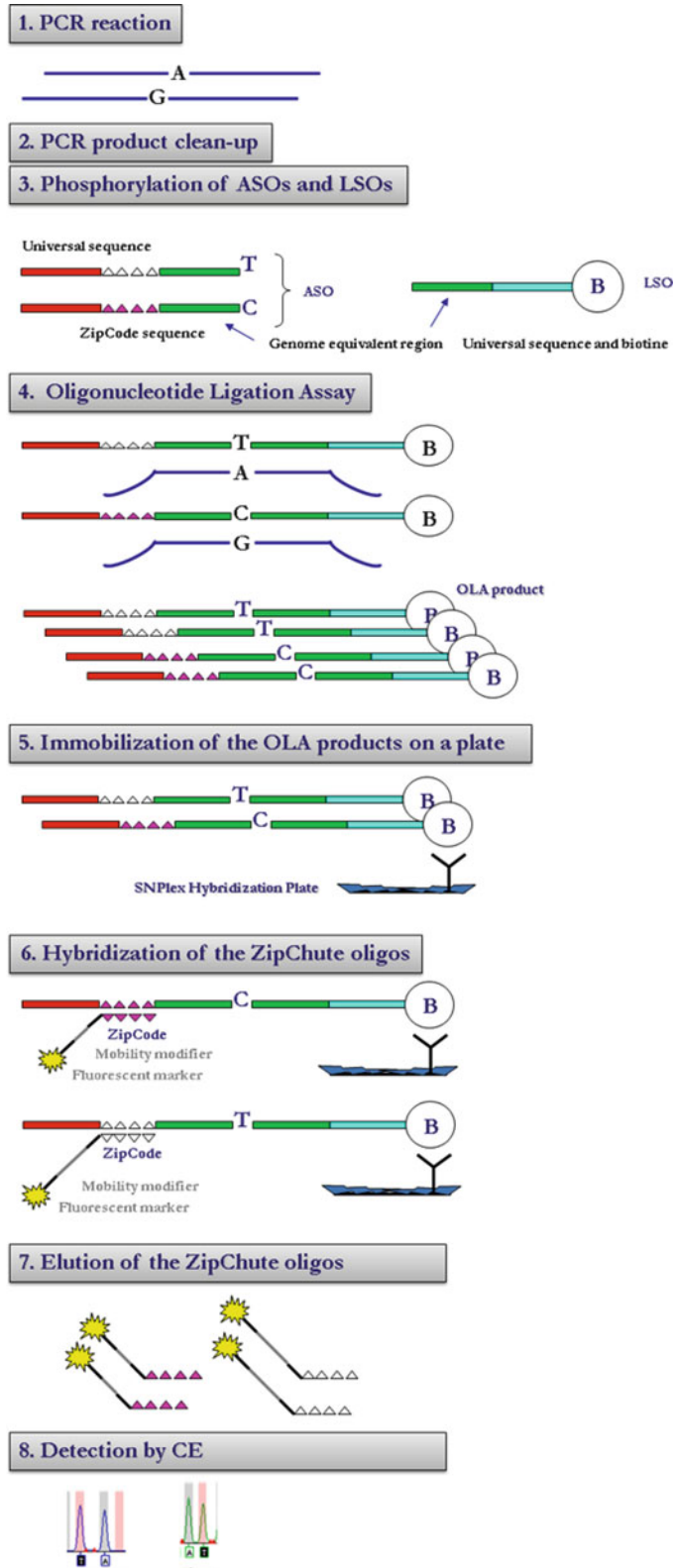
Fig. 1. General overview of the GenPlex™ protocol.
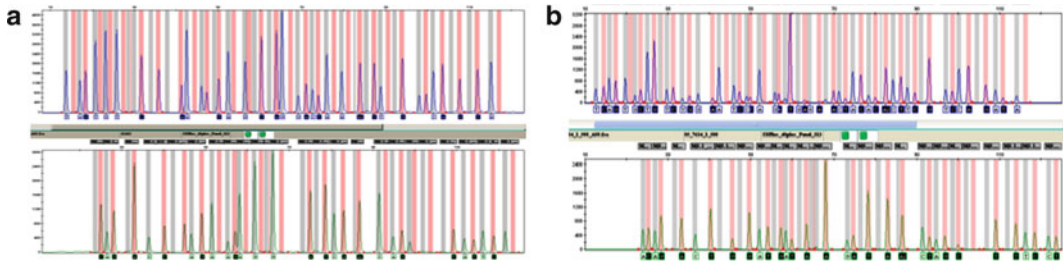
Fig. 2. Electropherograms of two samples analyzed with the GenPlex™ HID System. (**a**) Reference sample; (**b**) Sample with partly degraded DNA.

3. Deionized water.

4. Hybridization Binding Buffer SNPlex™ System (catalog number 4349304, Applied Biosystems).

5. Positive Hybridization Controls SNPlex™ System (catalog number 4349361, Applied Biosystems). Store at –15 to –25°C.

2.5. Hybridization of the ZipChute™ Oligos

1. Denaturant SNPlex™ System (catalog number 4349361, Applied Biosystems). Store at –15 to –25°C.

2. ZipChute™ Mix, 48-plex SNPlex™ System (catalog number 4349361, Applied Biosystems). Store at –15 to –25°C. Protect from light.

3. ZipChute™ Dilution Buffer SNPlex™ System (catalog number 4349306, Applied Biosystems).

4. Hybridization Wash Buffer (10×) SNPlex™ System (catalog number 4349301, Applied Biosystems).

2.6. Elution of the ZipChute™ Oligos

1. Hybridization Wash Buffer (10×) SNPlex™ System (catalog number 4349301, Applied Biosystems).

2. Sample Loading Reagent SNPlex™ System (catalog number 4349351, Applied Biosystems). Store at –15 to –25°C.

3. Size Standard, 48-plex SNPlex™ System (catalog number 4349351, Applied Biosystems). Store at –15 to –25°C. Protect from light.

2.7. Capillary Electrophoresis

1. Allelic Ladder, 48-plex SNPlex™ System (catalog number 4349351, Applied Biosystems). Store at –15 to –25°C. Protect from light.

2. 96-well Plate Septa and accessories (catalog number 4315933, Applied Biosystems).

3. 36- or 50-cm capillary array (catalog number depending on the DNA Analyzer and the number of capillaries, Applied Biosystems).

4. SNPlex™ System Array Conditioning Kit (catalog number 4352018, Applied Biosystems).

5. 10× Running Buffer with EDTA (catalog number 4335613, Applied Biosystems).

6. POP-7™ polymer for 3730/3730xl (catalog number 4363929, Applied Biosystems) or for 3130xl (catalog number 4352759, Applied Biosystems). Store at 2–8°C.

7. DS-40 Spectral Calibration Standard Kit (Dye Set S) (catalog number 4349365, Applied Biosystems). Store at –15 to –25°C. Protect from light.

8. Applied Biosystems 3730/3730xl or 3130xl DNA Analyzer.

9. Data Collection Software v2.0 (Applied Biosystems) or higher and necessary files associated: run module (HTSNP36_POP7; HTSNP50_POP7) and Dye Set S (S.zip).

2.8. Data Analysis

1. GeneMapper® Software v4.0 (Applied Biosystems).

2. GeneMapper® data analysis files for 3730/3730xl or 3130xl platforms (available from Applied Biosystems website):

    (a) Analysis method: SNPlex_Rules_3130.xml or SNPlex_Rules_3730.xml; GenPlex_NoCluster.

    (b) Bins and panels: SNPlex_48plex_3130_Bins.txt and SNPlex_48plex_3130_Panels.txt or SNPlex_48plex_3730_Bins.txt and SNPlex_48plex_3730_Panels.txt.

    (c) Assay Information File (AIF): SNP_for_ID_48.xlm.

    (d) Size standard: SNPlex_48plex_v1.xlm.

3. Methods

The GenPlex™ HID system protocol includes a large number of pipetting steps. It is recommended to use a pipetting robot when a medium or large number of samples are processed (see Note 1). Before starting with the protocol, design the sample plate layout. Include, together with the samples, a positive and a negative control per plate and one allelic ladder per injection. The location of the allelic ladder will depend on the Genetic Analyzer used to run the electrophoresis. The GenPlex™ protocol contains a number of steps where the plate layout needs to be taken into account. An overview of the procedure is shown in Fig. 1.

3.1. PCR Reaction

It is recommended to work on ice while setting up the PCR reaction.

1. Thaw the GenPlex™ Multiplex PCR Primer Mix and the QIAGEN Multiplex PCR Kit. Vortex the tubes and spin them briefly (see Note 2).

2. Prepare the master mix by adding 5 µL of Qiagen Multiplex PCR Kit, 2.5 µL of GenPlex™ Multiplex PCR Primer Mix and

0.5 μL of nuclease-free water. Scale the volumes to the desired number of samples (see Note 3).

3. Dispense 8 μL/well of master mix into a labeled MicroAmp® Optical 96-Well Reaction Plate. Leave the wells for the allelic ladder empty.

4. Add 2 μL of DNA (0.5–2 ng) to each well with master mix. Add 2 μL nuclease-free water to the well for the negative control.

5. Seal the plate with MicroAmp® Clear Adhesive Film. Make sure that the plate is completely sealed in order to avoid evaporation of the master mix during the PCR reaction (see Note 4). Vortex the plate and spin it briefly at a speed around 800 rpm.

6. Place the reaction plate in the thermal cycler (see Note 5).

7. Cover the plate with the thermal cycler heated lid. Set the reaction volume to 10 μL and use the following thermal cycling conditions: Denaturation at 95°C for 15 min; 30 cycles of 94°C for 15 s, 57°C for 30 s, and 72°C for 30 s; hold at 72°C for 10 min; hold at 4°C.

8. When the PCR run is complete, remove the plate from the thermal cycler. The plate may be stored at 4°C for a few days or at –20°C for several weeks.

3.2. PCR Product Clean-up with ExoSAP

1. Spin the PCR plate briefly before removing the adhesive film. Remove the film carefully in order to avoid cross-contamination. Add 2 μL of ExoSAP-IT® to each well.

2. Seal the plate with a new MicroAmp® Clear Adhesive Film. Vortex the plate briefly to mix the contents and spin the plate.

3. Place the reaction plate in the thermal cycler. Cover the plate, set the reaction volume to 12 μL and run the following program: hold at 37°C for 30 min; hold at 99.9°C for 30 min and hold at 4°C.

4. When the run is complete, remove the plate from the thermal cycler. The plate may be stored at 4°C during few days or at –20°C for several weeks.

3.3. Phosphorylation of the OLA Oligos (see Note 6)

1. Thaw the OLA Oligo Mix, the OLA Master Mix, and the dATP. Vortex the tubes and spin them shortly (see Note 2).

2. Mix 10 μL of OLA Master Mix, 0.1 μL of dATP, and 0.1 μL of OLA Oligo Mix. Scale the volumes to the desired number of samples. Vortex the tube and spin it briefly.

3. Dispense 10 μL/well of mix into a Microamp® Optical 96-Well Reaction Plate (see Note 7). Follow the same plate layout as the one used for the PCR reaction.

4. Seal the plate with MicroAmp® Clear Adhesive Film. Make sure that the plate is completely sealed, vortex the plate, and spin it briefly.

5. Place the reaction plate in a thermal cycler. Cover the plate, set the reaction volume to 10 µL and run the following program: Hold at 48°C for 30 min; hold at 90°C for 20 min; hold at 4°C.

6. When the run is complete, remove the plate from the thermal cycler. It is recommended to immediately proceed with the next step.

3.4. OLA Reaction

1. Spin the reaction plates from Subheading 3.2 (ExoSAP) and 3.3 (phosphorylation). Remove the films carefully and transfer 10 µL of phosphorylated OLA oligos into the plate containing ExoSAP-treated PCR products by following the plate layout. Discard the phosphorylation plate.

2. Seal the plate containing ExoSAP-treated PCR products and phosphorylated OLA oligos with a clean MicroAmp® Clear Adhesive Film. Make sure that the plate is completely sealed in order to avoid evaporation. Vortex the plate and spin it shortly.

3. Place the reaction plate in the thermal cycler. Set the reaction volume to 22 µL and run the following program: 25 cycles of 94°C for 15 s and 54°C for 30 s; hold at 99°C for 10 min; hold at 4°C.

4. After the run is complete, spin the plate briefly. The OLA products may be stored at –20°C.

3.5. Immobilization of the OLA Products on the Hybridization Plate

1. Dilute the Hybridization Wash Buffer (10×) 10 times with deionized water (1:9).

2. Thaw the Positive Hybridization Control (PHC; see Note 8). Vortex and quick-spin the tube.

3. Mix 25 µL of Hybridization Binding Buffer with 0.0125 µL of PHC in a 5 mL tube. Scale the volumes to the desired number of samples. Vortex the mix.

4. Wash the wells of the SNPlex™ Hybridization Plate 3 times by adding and removing 200 µL Hybridization Wash Buffer (1×) per well (see Note 9). Make sure that the hybridization plate is empty before proceeding with the next step.

5. Dispense 25 µL Binding Buffer mixed with PHC into each well of the hybridization plate. Follow the sample plate layout of the experiment. Leave the wells for the allelic ladder empty.

6. Transfer 5–10 µL (see Note 10) of OLA product to the wells containing Hybridization Binding Buffer and PHC and seal

the hybridization plate with adhesive film. The remaining OLA products may be stored at –20°C.

7. Incubate the hybridization plate at room temperature for 60 min on a shaking rotary shaker (at a speed around 700 rpm).

8. Spin the plate briefly, remove the film carefully, and remove the supernatant. The supernatant can be removed by inverting the plate on a stack of paper towels (see Note 11).

3.6. Hybridization of the ZipChute™ Oligos

1. Set the temperature of the rotary shaker (or the oven) to 37°C.

2. Wash the wells of the SNPlex™ Hybridization Plate 4 times with 200 μL of Wash Buffer (1×). Make sure that the wells are completely empty before proceeding with the next step.

3. Thaw the denaturant and ZipChute™ Mix. Vortex gently the ZipChute™ Mix and spin the tube.

4. Mix 11.25 μL denaturant, 13.7 μL ZipChute™ Dilution Buffer, and 0.05 μL ZipChute™ Mix. Scale the volumes to the desired number of samples. Vortex and spin the mix.

5. Dispense 25 μL of the hybridization master mix into each well of the SNPlex™ Hybridization Plate according to the plate layout. Leave the wells for the allelic ladder empty.

6. Cover the plate with Aluminum PCR Sealing film. Incubate the plate for 60 min at 37°C on a shaking rotary shaker (at a speed around 700 rpm). Protect the plate from the light during the incubation.

3.7. Elution of the ZipChute™ Oligos

1. Spin the hybridization plate briefly, remove the film carefully, and remove the supernatant.

2. Wash the wells of the hybridization plate 5 times with 200 μL of Wash Buffer (1×).

3. Invert the plate onto a stack of paper towels and centrifuge at $1600 \times g$ for 30 s.

4. Thaw the Sample Loading reagent and the size standard. Vortex the size standard gently.

5. Mix 16.9 μL Sample Loading reagent and 0.6 μL size standard. Scale the volumes according to the number of samples and allelic ladders.

6. Dispense 17.5 μL of Sample Loading reagent mixed with size standard into each well of the hybridization plate. Follow the plate layout and let the wells for the allelic ladder empty.

7. Cover the hybridization plate with Aluminum PCR Sealing film and incubate at 37°C for 5 min.

3.8. Capillary
Electrophoresis

Before running the first GenPlex™ electrophoresis plate, (1) import the run modules into the Data Collection Software; (2) install Dye Set S in the DNA Analyzer and perform a spectral calibration using the SNPlex™ Matrix Standard DS-40; (3) create an instrument protocol by selecting the imported run module and Dye Set S; (4) eventually, precondition the array by using the SNPlex™ System Array Conditioning Kit, if a new array is used (see Note 12).

1. Spin the hybridization plate briefly and remove the aluminum film carefully.

2. Prepare a 96-well electrophoresis plate according to the protocol corresponding to the DNA analyzer.

3. Transfer 10–17.5 μL (depending on the signal strength) of the Sample Loading Mix from the hybridization plate into the 96-well electrophoresis plate (except for the allelic ladder wells), according to the plate layout. Discard the hybridization plate.

4. Pipette 15 μL/well of Sample Loading reagent mixed with size standard to the wells for the allelic ladder. Add 1.25 μL/well of allelic ladder to these wells.

5. Assemble the electrophoresis plate and load it on the 3730/3730xl or 3130xl DNA Analyzer.

6. Run the plate using the adequate instrument protocol.

3.9. Data Analysis

Before starting with the data analysis, GeneMapper® v.4.0 must be configured for the GenPlex™ data analysis by importing the relevant data analysis files.

1. Export the FSA files to the computer where GeneMapper® v.4.0 is installed.

2. Open a new SNPlex™ project and add the sample files.

3. Identify the allelic ladders in the Sample Type column.

4. Select the appropriate Analysis Method (SNPlex_Rules_3130, SNPlex_Rules_3730, or GenPlex_NoCluster). A clustering method (see Fig. 3) or a nonclustering method can be chosen (see Note 13).

5. Select the appropriate Panel, SNPlex_48plex_Panel_3130, or SNPlex_48plex_Panel_3730, in the Panel column.

6. Select SNPlex_48plex_v1 in the Size Standard column.

7. Select SNP_for_ID_48 in the SNP set column.

8. Analyze the data (click the green arrow).

9. Review the quality of the size standard.

10. Review the data and export the genotypes as a table. If the data is analyzed using the nonclustering method, peak height ratios can be analyzed manually by exporting the data to a spreadsheet software application (see Note 13).
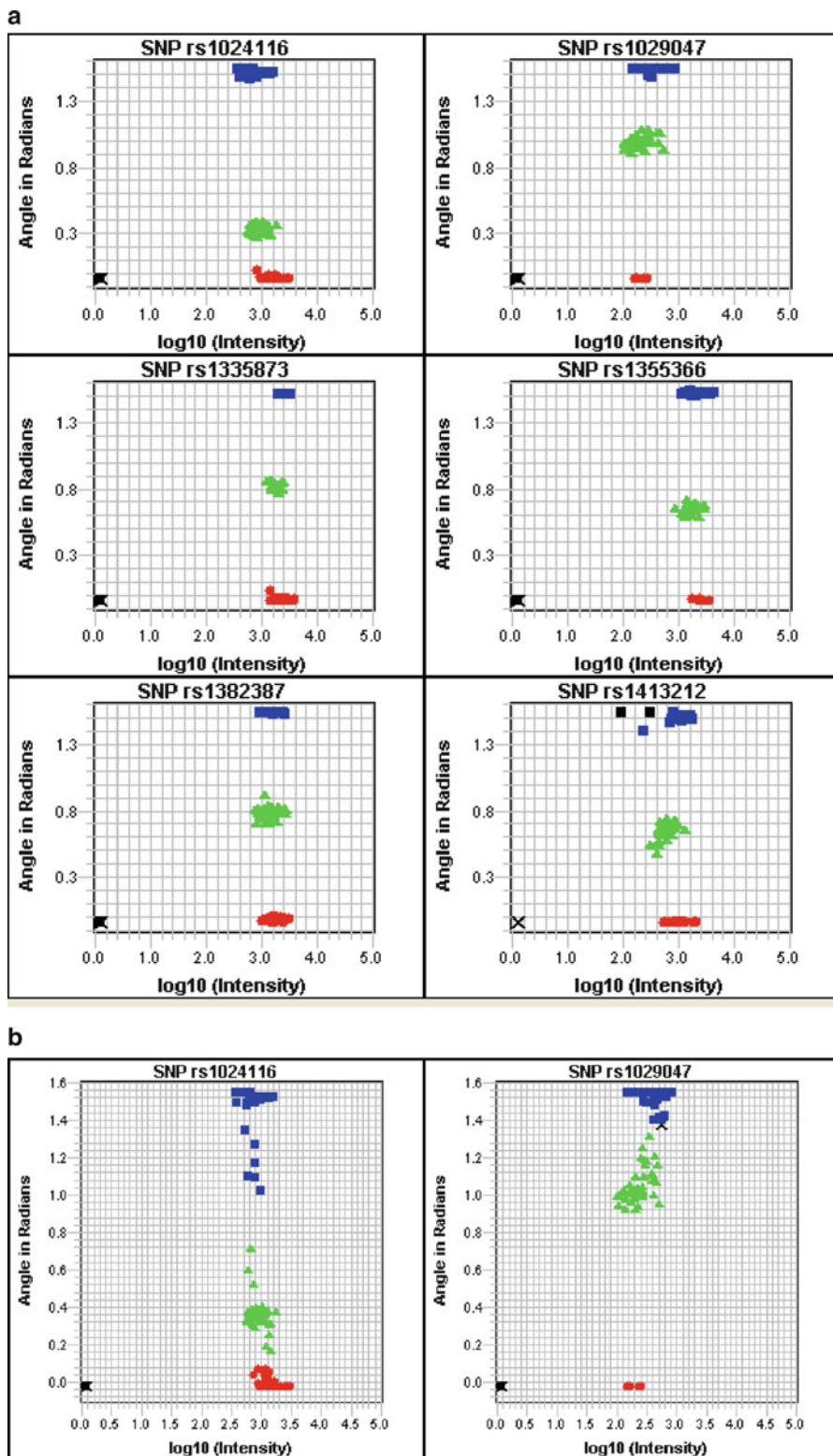
Fig. 3. GeneMapper® Cluster analysis. *Blue squares*: homozygotes of allele 1; *red circles*: homozygotes of allele 2; *green triangles*: heterozygotes. (**a**) Cluster analysis performed with reference samples. In this case, the clusters for homozygotes and heterozygotes are well-defined. (**b**) Cluster analysis performed including samples containing small amounts of DNA. Note the intermediate position of some of the samples between homozygotes and heterozygotes.

4. Notes

1. A detailed description of the use of a Biomek-3000 for the GenPlex™ protocol can be found in reference (19).

2. It is important to completely thaw the reagents in order to keep the expected concentration of the components. Vortex the primer and oligo mixes gently.

3. When scaling the number of samples, prepare extra volume in order to compensate for pipetting losses.

4. The efficiency of the reaction will be reduced if the master mix evaporates during the PCR reaction. In order to avoid evaporation, e.g., an ABI Prism® Optical Cover Compression Pad (catalog number 4312639, Applied Biosystems) can be used.

5. The protocol was optimized for a GeneAmp® PCR System 9700 thermal cycler with gold block (Applied Biosystems). The use of other thermal cyclers will probably require a reoptimization of the protocol.

6. Time can be saved by combining Subheadings 3.3 and 3.4 (phosphorylation and OLA reaction) in a single step. Nevertheless, better results were observed when two steps were performed.

7. Instead of dispensing the mix in a 96-well plate, the phosphorylation step can be performed in a 1.5 mL Eppendorf tube. A heating block may be used at 48 and 90°C.

8. The use of a positive hybridization control will help if troubleshooting is needed. If the only peak observed in the electropherogram is that corresponding to the PHC, it indicates that the hybridization did work and that the PCR and/or OLA reactions need to be optimized.

9. The exact volume of Wash Buffer used to wash the wells is not crucial. In the protocol defined for the Biomek-3000, a volume of 140 µL of Wash Buffer per well was used.

10. The volume of OLA product needed will depend on the amount of PCR products. If more than 500 pg of DNA are used in the PCR reaction, a volume of 5 µL of OLA product will most probably be sufficient.

11. It is recommended to perform the GenPlex™ procedure in two separated areas (1) a pre-PCR area (see Subheading 3.1) and (2) a post-PCR area (see Subheadings 3.2–3.8). There are many steps where the tubes containing PCR and OLA products are opened and handled, therefore extreme care is required in order to avoid contaminations.

12. The resolution of the CE machine can be tested by running a diluted solution of the SNPlex ZipChute™ Mix (1:224) and the internal size standard.

13. The peak height ratio of two alleles in heterozygotes varies depending on the SNP, the amount of DNA used in the PCR reaction and the electrophoretic conditions. The same is applicable for peak high ratios in homozygotes (peak ratio between the real allele and any spurious peak). Each SNP behaves differently under different conditions and criteria for allele calling must be established by each laboratory by calculating independent cut off values for homozygotes and heterozygotes. For this reason, it is highly recommendable to run a set of samples (around 100 samples) that can be used as reference for future experiments. This set of reference samples can be both used with the Cluster and noncluster methods. If the Cluster method is chosen, import the reference samples to the project containing the samples of the experiment. GeneMapper® will use all the samples to define clusters for homozygotes and heterozygotes.

If the noncluster method is used, export the peak heights of the reference samples to a spreadsheet software application and calculate peak height ratios and cut off values for homozygotes and heterozygotes. Export the peak heights of the samples of the experiment to a spreadsheet software application, calculate peak height ratios and evaluate them according to the cut off values calculated from the reference samples. Take into account that it may be necessary to update the set of reference samples if changes in the protocol are introduced or a new CE machine is used.

It is not recommendable to use the Cluster method if poor quality samples are analyzed (i.e., when less than 100 pg of DNA are used in the PCR reaction). When low amounts of DNA are used, peak height ratios change in an unpredictable way and some of the SNPs can be miscalled. An example is shown in Fig. 3b where a small amount of DNA was used in the PCR reaction and no clear clusters could be defined for homo- and heterozygotes. A manual scrutiny of the peak height ratios using a spreadsheet software application is recommended when undefined clusters are observed. As a general rule, SNP types that show unexpected peak height ratios should be failed.

Unusual peak height ratios can also be observed in samples wwwwith DNA mixtures. Results with a high overall signal (high RFUs) and unusual peak height rations most probably indicate the presence of a DNA mixture in the sample. Unusual peak height rations in low signal samples may be due to the use of small amounts of DNA in the PCR reaction.

References

1. Børsting C, Sanchez JJ, Hansen HE et al (2008) Performance of the SNP*for*ID 52-plex assay in paternity testing. Forensic Sci Int Genetics 2:292–300

2. Børsting C, Rockenbauer E, Morling N (2009) Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard. Forensic Sci Int Genetics 4:34–42

3. Krjutškov K, Viltrop T, Palta P et al (2009) Evaluation of the 124-plex SNP typing microarray for forensic testing. Forensic Sci Int Genet 4:43–48

4. Brión M, Sanchez JJ, Balogh K et al (2005) Introduction of a single nucleodite polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages. Electrophoresis 26(23):4411–4420

5. Tomas C, Sanchez JJ, Castro JA et al (2010) Forensic usefulness of a 25 X-chromosome single-nucleotide polymorphism marker set. Transfusion 50(10):2258–2265

6. Mengel-From J, Børsting C, Sanchez JJ et al (2010) Human eye colour and HERC2, OCA2 and MATP. Forensic Sci Int Genet 4(5):323–328

7. Sanchez JJ, Phillips C, Børsting C et al (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27:1713–1724

8. Sanchez JJ, Børsting C, Balogh K et al (2008) Forensic typing of autosomal SNPs with a 29 SNP-multiplex – Results of a collaborative EDNAP exercise. Forensic Sci Int Genetics 2:176–183

9. Musgrave-Brown E, Ballard D, Balogh K et al (2007) Forensic validation of the SNP*for*ID 52-plex assay. Forensic Sci Int Genetics 1:186–190

10. Mengel-Jørgensen J, Sanchez JJ, Børsting C et al (2005) Typing of Multiple SNPs Using Ribonuclease Cleavage of DNA/RNA Chimeric Single Base Extension Primers and Detection by MALDI-TOF Mass Spectrometry. Anal Chem 77:5229–5235

11. Børsting C, Sanchez JJ, Morling N (2005) SNP typing on the NanoChip electronic microarray. Methods Mol Biol 297:155–168

12. Harrison C, Musgrave-Brown E, Bender K et al (2006) A sensitive issue: Pyrosequencing as a valuable forensic SNP typing platform. International Congress Series 1288:52–54

13. Sobrino B, Lareu M, Brion M et al (2004) SNP genotyping with single base extension-tag microarrays. International Congress Series 1261:331–333

14. AB SNPlex™ chemistry guide pdf: http://docs.appliedbiosystems.com/pebiodocs/04360856.pdf

15. The SNP*for*ID consortium: htpp://www.snpforid.org

16. Phillips C, Fang R, Ballard D et al (2007) Evaluation of the Genplex SNP typing systems and a 49plex forensic marker panel. Forensic Sci Int Genetics 1:180–185

17. Tomas C, Stangegaard M, Børsting C et al (2008) Typing of 48 autosomal SNPs and amelogenin with GenPlex SNP genotyping system in forensic genetics. Forensic Sci Int Genetics 3:1–6

18. Tomas C, Axler-DiPerte G, Budimlija ZM et al (2011) Autosomal SNP typing of forensic samples with the GenPlex™ HID System: Results of a collaborative study. Forensic Sci Int Genetics 5:369–375

19. Stangegaard M, Tomas C, Hansen AJ et al (2008) Biomek-3000 and GenPlex SNP genotyping in forensic genetics. JALA 13:297–303

Chapter 7

Typing of 49 Autosomal SNPs by Single Base Extension and Capillary Electrophoresis for Forensic Genetic Testing

Claus Børsting, Carmen Tomas, and Niels Morling

Abstract

We describe a method for simultaneous amplification of 49 autosomal single nucleotide polymorphisms (SNPs) by multiplex PCR and detection of the SNP alleles by single base extension (SBE) and capillary electrophoresis. All the SNPs may be amplified from only 100 pg of genomic DNA and the length of the amplicons range from 65 to 115 bp. The high sensitivity and the short amplicon sizes make the assay very suitable for typing of degraded DNA samples, and the low mutation rate of SNPs makes the assay very useful for relationship testing. Combined, these advantages make the assay well suited for disaster victim identifications, where the DNA from the victims may be highly degraded and the victims are identified via investigation of their relatives. The assay was validated according to the ISO 17025 standard and used for routine case work in our laboratory.

Key words: SNP typing, Multiplex PCR, Single base extension, Capillary electrophoresis, Forensic genetics, SNP*for*ID, ISO 17025 standard, Accreditation

1. Introduction

There are five main reasons why single nucleotide polymorphisms (SNPs) may be preferred to the standard short tandem repeats (STRs) loci used in most forensic genetic investigations. First, SNPs may be amplified on short amplicons, which increase the chance of obtaining results from highly degraded DNA and make the amplification of the loci very efficient. Second, SNPs have low mutation rates (approximately 10^{-8}), which is a major advantage in relationship testing. A mutation will often lead to a genotype that is incompatible with Mendel's law of inheritance and may result in comparable likelihoods for different family scenarios. Third, most SNPs are bi-allelic, and thus, relatively few database samples

are required to obtain reasonable estimates of the allele frequencies in a population. Fourth, SNPs may become an important investigative tool for the police when SNP packages for physical characteristics or ethnic origin are fully developed and validated. Fifth, the highly automated, high-throughput SNP genotyping technologies are interesting for the forensic community, because of the dramatic increase in sample numbers and the demand for faster and cheaper forensic genetic investigations that many forensic laboratories have experienced in recent years.

The SNP*for*ID consortium (www.snpforid.org) developed a SNP typing assay for human identification purposes that made it possible to amplify 52 SNPs from diminutive amounts of genomic DNA (1). All SNP loci were amplified in one PCR and the SNPs were analysed by two single base extension (SBE) reactions. In the SBE reaction, the SBE primers hybridized to the PCR products and were extended with a labelled ddNTP complimentary to the nucleotide in the SNP position. The SBE products were analysed by capillary electrophoresis, where the length of the extended SBE primer identified the SNP locus and the colour of the ddNTP label identified the SNP allele. The two SBE reactions were injected into the same capillary with 5 min intervals and analysed in one data file (1, 2).

The main criteria for selection of the SNP loci in the SNP*for*ID multiplex were high levels of heterogeneity in European, Asian and African population groups and a minimum distance of 100 kb between the SNPs and neighbouring genes. The SNPs were shown to be polymorphic in populations from four different continents and there was not found any evidence of linkage disequilibrium for any pair of SNPs in the multiplex (1).

The sensitivity and the overall robustness of the SNP*for*ID multiplex assay were improved when the assay was validated for forensic genetic case work according to the ISO 17025 standard. In the validation process, the number of SNPs typed with the assay was reduced from 52 to 49 to allow optimization of the SBE primer design and simplification of the analysis of the electropherograms (2). With the improved assay, full SNP profiles were obtained from only 100 pg of genomic DNA (3). The length of the PCR amplicons in the SNP*for*ID multiplex ranged from 65 to 115 bp and full SNP profiles were generated from highly degraded DNA samples (Fig. 1), where it was not possible to amplify any or just a few STR loci with the commercial STR kits ((1, 4), unpublished results). The mean match probability ranged from 10^{-14} to 10^{-21} and the paternity indices ranged from 10^5 to 10^6 for trios (1, 2, 5–9). More than 75 population groups have been typed for these SNPs since the assay was developed and the allele frequencies are freely available at the SNP*for*ID homepage (10).

Here, we describe the SNP typing protocols used for routine case work in our laboratory. The SNP*for*ID multiplex is used in
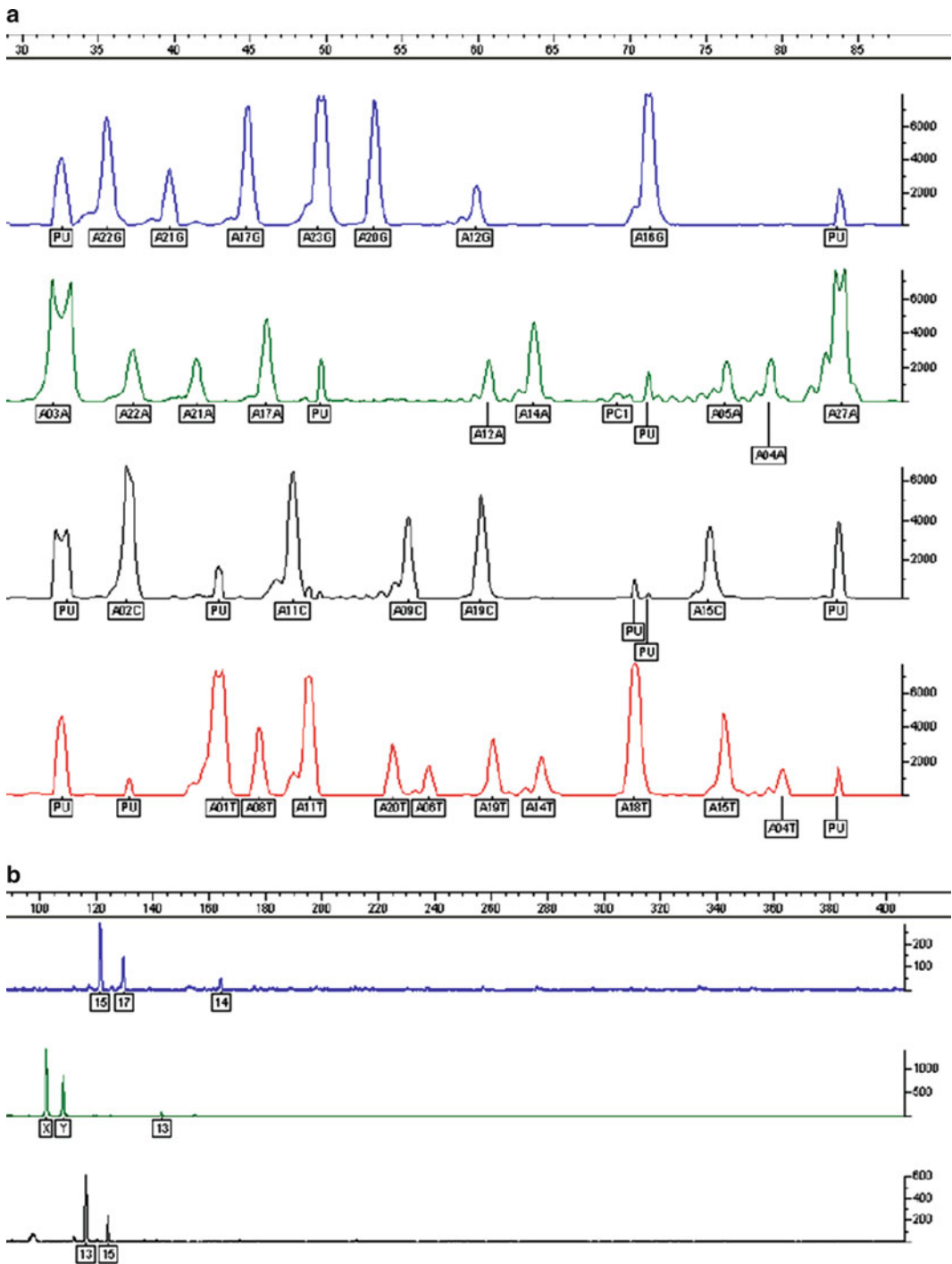
Fig. 1. A corpse was washed up on a beach after an unknown period of time in the water and DNA was extracted from a nail. The sample was typed with the SNP*for*ID multiplex (**a**) and the AmpF*l*STR® SEfiler plus™ PCR amplification kit (Applied Biosystems) (**b**). The same amount of DNA was used in the two PCRs. (**a**) The results from SBE reaction A (see Subheading 3.3) with 21 of the 49 SNPs are shown. Allele calls are shown in boxes, e.g. "A22G", which is the G allele of SNP number A22 (rs733164). "PU" is a pull-up and "PC1" is a ddATP-extended PCR fragment. A full SNP profile was obtained from the sample. The match probability was 1.6×10^{-21}. (**b**) Amelogenin and two STR loci (D3S1358 and D19S433) were typed with the AmpF*l*STR® SEfiler plus™ PCR amplification kit. Possible allele drop-outs were observed at two other STR loci (D8S1179 and vWA). Allele calls are shown in boxes. No PCR fragments longer than 164 basepairs were observed and no fragments with a PET® (*red*) dye were detected. The match probability based on the results from D3S1358 and D19S433 was 9.0×10^{-3}.

relationship cases where the standard typing of 15 STRs does not provide sufficient information or leads to an ambiguous conclusion.

2. Materials

2.1. Multiplex PCR

1. 5 U/μL AmpliTaq Gold® DNA polymerase (Applied Biosystems (AB), Carlsbad, USA). The enzyme must be stored at –20°C (see Note 1).
2. 10× GeneAmp PCR buffer I with 15 mM MgCl$_2$ (AB, Carlsbad, USA). The buffer must be stored at –20°C.
3. 25 mM MgCl$_2$ (AB, Carlsbad, USA). The solution must be stored at –20°C.
4. 10 mM dNTP mix (GE healthcare, Pollards Wood, UK). The solution must be stored at –20°C (see Note 2).
5. High-performance liquid chromatography (HPLC) purified oligonucleotides (DNA Technology, Aarhus, Denmark) (see Note 3). Prepare the PCR primer mix according to Table 1. The primer mix must be stored at –20°C (see Note 2).
6. Purified human genomic DNA (see Note 4).

2.2. Degradation of dNTPs and PCR primers

1. ExoI-SAP solution. Mix 2 μL 10 U/μL Exonuclease I (GE healthcare, Pollards Wood, UK) and 50 μL 1 U/μL Shrimp Alkaline Phosphatase (SAP) (GE healthcare, Pollards Wood, UK). The enzymes must be stored at –20°C (see Note 5).

2.3. Multiplex SBE

1. ABI Prism SNaPshot™ Multiplex Kit (AB, Carlsbad, USA) with DNA polymerase, buffer and labelled ddNTPs. The kit must be stored at –20°C. The ABI Prism SNaPshot™ Multiplex Kit is sensitive to light and the tube should be wrapped in aluminium foil when it is outside the freezer.
2. High-performance liquid chromatography (HPLC) purified oligonucleotides (DNA Technology, Aarhus, Denmark) (see Note 3). Prepare the SBE primer mix A and B according to Tables 2 and 3, respectively. The primer mixes must be stored at –20°C (see Note 2).
3. Ammonium sulphate (Sigma).

2.4. Degradation of ddNTPs

1. 1-U/μL Shrimp Alkaline Phosphatase (GE healthcare, Pollards Wood, UK). The enzyme must be stored at –20°C.

2.5. Capillary Electrophoresis

1. GeneScan™: 120 LIZ™ size standard (AB, Carlsbad, USA). The size standard must be stored at 2–8°C.
2. Hi-Di™ formamide (AB, Carlsbad, USA). The formamide must be stored at –20°C.

Table 1
PCR primer mix

Marker code	NCBI number	PCR forward primers	PCR reverse primers	Concentration (µM) in primer mix	Amplicon size (bp)
A01	rs1490413	GTGTGGACTGGGCTGATGT	TTCTCACTAGTGTCCCTGCTCTG	0.07	68
A02	rs876724	GCAGGCTCCATTTTATACCACT	GAATATCTATGAGCAGGCAGTTAGC	0.05	83
A03	rs1357617	AGCTGATGCAGACCACTTCAC	GGATAGCTGATAAGAAACATGACCA	0.05	90
A04	rs2046361	CCTATTTGTATGTATCTATTGTCTATGAACG	GTCATTGTTGACACTTCACCTTCTA	0.14	79
A05	rs717302	CTTTAGAAAGGCATATCGTATTAACTGTG	AACACAGAAAGAGGTTCATATGTTGG	0.07	86
A06	rs1029047	CATAACGTGGATTTGTCAGCA	GGAATAAACTGAAGGCTAAAGAAAAG	0.18	100
A07	rs917118	GCCCTTTAGGGTCGGTTC	GTAAGAGATGACTGAGGTCAACGAG	0.08	87
A08	rs763869	ATCAAGTGCTTTCTGTTGACATTTG	GGCTACTCCCTCATAATGTAATGC	0.18	100
A09	rs1015250	AAGTGATGGAGTTAGGAAAAGAACC	AAGACATTAGGTGGATTCATAGCTG	0.09	95
A10	rs735155	GGAGAAAACCGGAGAGCTG	GAGTGTCACCGAATTCAACG	0.08	100
A11	rs901398	CTGGGTGCAAACTAGCTGAATATC	CTGGAATGTACTAGGCAAGAAACTAA	0.06	70
A12	rs2107612	GAGCATTCTCTTCTGTTAAAATTGC	TGAGTACATTATTCAACTGTTTTGGAG	0.09	93

Table 1
(continued)

Marker code	NCBI number	PCR forward primers	PCR reverse primers	Concentration (µM) in primer mix	Amplicon size (bp)
A14	rs1454361	AGGGAAATACACCCTGAGCTG	AGCTGTCCATCATCAGT AAGACAC	0.05	73
A15	rs2016276	TGCATCCCAGCCTCCACT	ATTGTACCTTGCCACTTT GTGTG	0.32	90
A16	rs729172	CATTAATATGACCAAGGCT CCTCT	ACATTTcCCTCTTGCGGTTAC*	0.05	60
A17	rs740910	GTATAACAGTTTGCTAAG TAAGGTGAGTG	AGATAGGTTCGAGTTTTG GCTTTA	0.15	87
A18	rs1493232	CTATTCTCTCTTTTGGG TGCTAGG	CAAACTGTTTATTGTGAG GCCTGT	0.13	59
A19	rs719366	CCACAGCATCTTTTAACTC TTTTATTATCC	GTAAGGACTTATAGTGAGTA AAGGACAGG	0.11	105
A20	rs1031825	CTTATCTTTCCCACATTAT GGTCCT	AAGATATAATCACTGCTTT CAAGTATGC	0.09	98
A21	rs722098	GGAAGTACACATCTGTTGA CAGTAATGA	GGGTAAAGAAATATTCAGC ACATCC	0.24	81
A22	rs733164	AGCTTTCAGCCCCAGGTC	CGGCTCAGGAATGTCAGG	0.07	68
A23	rs826472	TGAATTTTGTCTCGTTTATAT TAGTCACC	TGTAATTGAAATTTGTAGGC AATAGAC	0.11	85
A24	rs2831700	GGCTAAACTATTGCCGGAGA	TTCCCTAGAACCACAATTA TCTGTC	0.04	62
A25	rs873196	GCATTCAAATCCCAAGTGCT	GCAGGAGTTGGAGTCA ATCAG	0.03	63

A26	rs1382387	ACGAAGGAGAAACAC CTGAACT	TGGAGTACTTAATAAGAC GCTGCAT	0.05	69
A27	rs2111980	AGCATCTTGGCAGCATCC	AGCAAGATCTTTGCCAG TGAGT	0.04	72
A28	rs2056277	CCAAACTGGGTGTTAGGGAGAC	TCATTATCTCGTCATACT TCCCTGT	0.04	73
A29	rs1024116	CCATGTGTTCTAATAAAA AGGATTGC	TGGGAAGTGAGCAAAAG TAAATACA	0.03	76
A30	rs727811	GTGTTTCTTTTTCTCTT ACCGGAAC	GTGAATGAAATCATGAG ATTGCTG	0.05	78
A32	rs1413212	AACCTCCTTTGGAAACACTGAC	CAACATTCCATTATCCA GGAGAC	0.06	84
A33	rs938283	CATTGAAGTCCTAACCCC TAGTACG	GGATGAGGCCCACCCATA	0.07	85
A34	rs1979255	TCAGAGACTATGGATGGT ATTTAGGTC	CATGGAACGTTGGAAC TCTTG	0.05	86
A36	rs2076848	GCCTCACCACCAGAAATCAG	GACATCAGAAACTCCC ATGAAACT	0.10	89
A37	rs1355366	CCATGATTTTCTTGTGGTGAGA	CACATGTGCTTAGGCCACAAC	0.10	90
A38	rs907100	GGAGTTCCTGATAACGAT TCTGAAG	ACAGAAAAGAAGCcCAGTTGGA*	0.23	91
A39	rs354439	GGCTTCTCTTTCCCTTA TGTATCTC	CAGGTTGCGATAGAAA ACAGTGAAT	0.06	93
A40	rs2040411	TCTGGAATGCCAGTTCTTTTGT	CAGAACGCCTATGAA AACCAGT	0.07	94
A41	rs737681	ACATGTGAGGCCATCTCCAC	CCTTACTGTGATGTAGG CACTGTTC	0.03	96

(continued)

Table 1
(continued)

Marker code	NCBI number	PCR forward primers	PCR reverse primers	Concentration (µM) in primer mix	Amplicon size (bp)
A42	rs2830795	CACTTCTATAGACATAGGACACACCAT	ATCTAGGCTCTGAATCAGGATGAG	0.04	97
A43	rs251934	AGAGGGCAGTGAGGCTTTTAAGTAG	TGCTAGAATCCAGACTTAACTACCAG	0.03	98
A44	rs914165	AGCAGCAGAGCCTGGATG	AGACCAGTCACCTCTTTTGCACT	0.10	100
A45	rs10495407	AGATCTCCACTTCCTCTTGGTTG	CTCCCAAATTTACATTGCCACT	0.06	102
A46	rs1360288	AGACTCTCTGTTGTGTGGCTTTG	GAGGGGGCATCTGTTGAG	0.09	103
A48	rs964681	GTACCTGGAGGTGATTTCTGTGAG	GTTATGGAGGATTGGTAAGAACCAG	0.08	106
A49	rs1005533	GGTTTGTGTGTGAGTGTTTCAGAT	CCTTATGCCTCCCCTGAAC	0.04	107
A50	rs8037429	TTCACTTTGCTACACCTCCATAGTA	TGCTACGTAAGAGGTCATTGCTATC	0.05	108
A51	rs891700	TTTTCAGAGGTGGTATTTCTAGCTG	GCTATGACACTCCTTAGAACTATGCAA	0.05	109
A52	rs1335873	GTGGATGATATGGTTTCTCAAGG	TTCAACAAACGTGTGATGCTCT	0.04	110
A54	rs1528460	TCCTGGAGATCAATATTTAGCCTTA	GGGTGACCAGTAGTTCTATGAGC	0.12	115

PCR primer sequences, primer concentrations and amplicon sizes. Bases shown in lower case are pyrimidines (C/T)

Table 2
SBE primer mix A

Marker code	NCBI number	SBE primer mix A	Concentration (µM) in primer mix	Detected alleles
A01	rs1490413	CTCTCTCTCTCTCTCTCTATCTGAGGCCAGCCAGTT	0.33	C/T
A02	rs876724	CTCTCTCTCTCAACACTGCACTGAAGTATAAGTA	0.58	C/T
A03	rs1357617	CTCTGACAATGATAAGAAACATGACCAAGC	0.83	A/T
A04	rs2046361	CCCCCCCCCCCCCCAACTGACTAAACTAGGTGCCACGTCGTGAAAG TCTGACAATTGTTGACACTTCACCTTCTA	0.50	A/T
A05	rs717302	CCCCCCCCCCCAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACA AGGCATATCGTATTAACTGTGT	0.17	G/A
A06	rs1029047	AAACTAGGTGCCACGTCGTGAAAGTCTGACAAGTAAGAATTCAAGATGGTATTT	1.00	A/T
A08	rs763869	CGTGAAAGTCTGACAAGTTTGTTTATATTATTTCTAACTCAA	1.17	C/T
A09	rs1015250	CTCTCTCTCTCTCTCTCTCTCTCTCTATGAAAAATCACTGGGACAG	0.67	G/C
A11	rs901398	AGGTGCCACGTCGTGAAAGTCTGACAAACTAGCTGAATATCAGCCC	0.33	C/T
A12	rs2107612	TAAACTAGGTGCCACGTCGTGAAAGTCTGACAAGTGTTTTTCTAAATCA TATTGTCT	0.33	G/A
A14	rs1454361	CCCAAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAAAAATACAC CCTGAGCTGC	0.50	A/T
A15	rs2016276	CCCCCCCCCCCCAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTG ACAATGAGAGAGAGAGAAAAATGTGC	0.33	C/T
A16	rs729172	CCCCCCCCCAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAATC TTGCGGTTACATACAITC	0.17	G/T
A17	rs740910	CGTCGTGAAAGTCTGACAATAAGGTGAGTGGTATAATCATAT	0.17	G/A

(continued)

Table 2
(continued)

Marker code	NCBI number	SBE primer mix A	Concentration (µM) in primer mix	Detected alleles
A18	rs1493232	CCCCCAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAATTGTGAG GCCTGTTTATTTTG	0.53	G/T
A19	rs719366	ACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAATTTCCTCCTCCCA TTCTAG	1.25	C/T
A20	rs1031825	AAACTAGGTGCCACGTCGTGAAAGTCTGACAATTTCAAGTATGCTCGGGG	0.83	G/A
A21	rs722098	CGTGAAAGTCTGACAATGACAGTAATGAAATATCCTTG	0.50	G/A
A22	rs733164	CGTGAAAGTCTGACAAGGCCATCCCACTTGGAAA	0.17	G/A
A23	rs826472	GTGCCACGTCGTGAAAGTCTGACAAGACTAATTAATCTGCTATCCTG	0.33	G/A
A27	rs2111980	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT CTCTCTCTCTCTCTCTCTCTCTCTCTCCATCTTGGCAGCATCCTTC	0.67	G/A

SBE primer sequences, primer concentrations and detected alleles. Add ammonium sulphate to a final concentration of 160 mM to minimize primer–dimer formation

Table 3
SBE primer mix B

Marker code	NCBI number	SBE primer mix B	Concentration (µM) in primer mix	Detected alleles
A07	rs917118	TGACTGAGGTCAACGAGC	1.01	G/A
A10	rs735155	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT CTCTCCACCGAATTCAACGGGAAG	0.21	C/T
A24	rs2831700	TCTCTCTCTCTCTCTCTCTCTCTCTCTTATCTGTCTTTGGC CTTCTAACT	0.76	C/T
A25	rs873196	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAAT CAGGCATGTTCACATTACAA	0.95	A/G
A26	rs1382387	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGAGAAA CACCTGAACTTTCAA	0.50	G/T
A28	rs2056277	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC TCTCTCTCTCTCTCTCTCTCTCTCCGTGTCTTGAACTCCC	0.13	A/G
A29	rs1024116	TCTCTCTCTCTCTCTCTCTCTCTCTCTTGTTCTAATAAAAAG GATTGCTCAT	0.76	A/G
A30	rs727811	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCGG AACTTCAACGACTTA	0.63	A/C
A32	rs1413212	CTCTCTCTCTCTCTCTCTAGACATTTGTTCATATAAGTGAAATG	0.63	C/T
A33	rs938283	TCTAGTACGTTAGATGTGACCG	0.45	C/T
A34	rs1979255	TCTCTCTCTCTCTCTCTCTCTGAACGTTGGAACTCTTGCT	0.50	C/G
A36	rs2076848	CTCTCTCTCTCCACCAGAAATCAGGGC	0.76	A/T
A37	rs1355366	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTGGCTTTGGCAAGTCC	0.38	C/T
A38	rs907100	TCTCTCTCTCTCCAGTTGGGAGCCTTCTTT	0.76	C/G

(continued)

Table 3
(continued)

Marker code	NCBI number	SBE primer mix B	Concentration (µM) in primer mix	Detected alleles
A39	rs354439	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC TCTCTCTCTCGCCTCGCTCGCTCATGTATCACATTCCT	0.19	A/T
A40	rs2040411	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCTCTG TATTTTCTTACTCTAAGTGC	0.25	A/G
A41	rs737681	TGAGGCCATCTCCACC	0.63	C/T
A42	rs2830795	TCTCTCTCTCTCTCTCTGGACACACCATTTATTGTCTAAAG	1.14	A/G
A43	rs251934	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT CTCTCTCTCTCTCTCTTAGTAGATATCTGGCTGTCCC	0.38	A/G
A44	rs914165	TCTCTCTCTCTCTCTGAGCTGCCTGAGGGTGT	1.01	C/T
A45	rs10495407	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGAAA AACAACAGAATGGAATAGGA	0.88	C/T
A46	rs1360288	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAGGG GATGCAGCCAGT	0.19	A/G
A48	rs964681	TCTCTCTCTCTCTCTCTTGTGAGGAACGTCGACA	0.50	C/T
A49	rs1005533	TCTCTCTCTCTCTCTCTGCAAGAGCCGTGGAATT	1.01	A/G
A50	rs8037429	TCTCTCTCTCTCTCTCTCTCTCTCTCTC TCTCTCTCTCTTACAGAATAACTCTTGCGCTAC	0.95	A/G
A51	rs891700	TCTCTCTCTCCAATTCTTTTTTTTTGAAGCCT	0.19	A/G
A52	rs1335873	TCTCTCTCTCTAGGTACCTAGCTATGTACTCAGTAT	0.44	A/T
A54	rs1528460	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT CTCTAAAACTTTTACTGTAACATTTTCA	1.01	A/G

SBE primer sequences, primer concentrations and detected alleles. Add ammonium sulphate to a final concentration of 160 mM to minimize primer–dimer formation

3. Performance Optimized Polymer 4™ (POP-4) (AB, Carlsbad, USA) (see Note 6). POP-4 must be stored at 2–8°C. At room temperature, e.g., on the capillary electrophoresis instrument, POP-4 is only stable for 1 week.

4. 3730 10× running buffer (AB, Carlsbad, USA). The buffer must be stored at 2–8°C. At room temperature, e.g., on the capillary electrophoresis instrument, the buffer is only stable for two days.

5. DS-02 (dye set E5) Matrix Standard with dR110, dR6G, dTAMRA™, dROX™ and LIZ® dyes (AB, Carlsbad, USA). The matrix standard must be stored at –20°C.

6. 3130xl/3100 Genetic Analyzer Capillary Array, 36 cm 16-capillary array (AB, Carlsbad, USA).

7. Capillary electrophoresis instrument, e.g. AB310, AB3100, AB3100 Avant, AB3130, AB3130XL or AB3500 genetic analyzers (AB, Carlsbad, USA).

2.6. Analysis of the SBE Products

1. GeneScan® 3.7.1 software (AB, Carlsbad, USA).

2. Genotyper™ 3.7 software (AB, Carlsbad, USA).

3. Microsoft® Excel® spreadsheet software (Microsoft Corporation, Redmond, USA).

3. Methods

The methods described below outline (1) the multiplex PCR, (2) degradation of the dNTPs and the PCR primers left over from the PCR, (3) the multiplex SBE reaction, (4) degradation of the ddNTPs left over from the SBE reaction, (5) capillary electrophoresis, and (6) analysis of the SBE products.

3.1. Multiplex PCR

PCRs should be mixed in dedicated pre-PCR work areas and preferable in benches with laminar air flow to avoid contamination (see Note 7). The protocol shown below was designed for FTA-card samples, where a 1.2-mm disk was punched out of the FTA-card into a 96-well PCR plate, washed with ddH$_2$O and dried (11, 12). However, liquid DNA preparations (see Note 4) may also be used with a few adjustments.

1. Thaw the reagents. Except for the AmpliTaq Gold® DNA polymerase, the reagents should be shaken briefly on a Vortex mixer and placed on ice. The AmpliTaq Gold® DNA polymerase should only be removed from the freezer in a container cooled down to –20°C.

2. Make a reaction mixture with 0.4 μL 5 U/μL AmpliTaq Gold® DNA polymerase, 2.5 μL 10× GeneAmp PCR buffer I, 6.5 μL

25 mM MgCl$_2$, 1.75 µL 10 mM dNTP mix, 13 µL PCR primer mix (Table 1) and 0.85 µL nuclease-free water for each sample. If the sample is a liquid preparation of DNA, no water should be added to the reaction mixture. The reaction mixture should be shaken briefly on a Vortex mixer and placed on ice.

3. Add 25 µL reaction mixture to each well in the PCR plate with a washed and dried FTA-card punch. If the sample is a liquid preparation of DNA, add 24 µL reaction mixture and 1–4 µL sample to each well. Include a positive (sample with known SNP profile) and a negative (water instead of DNA) sample control.

4. Seal the PCR plate with adhesive seal and spin it briefly in a plate centrifuge at 600 rpm.

5. Place the PCR plate in a PCR thermocycler (see Note 1) and run the following PCR cycle programme: Denaturation at 95°C for 10 min followed by 35 cycles of 95°C for 30s, 60°C for 30 s and 65°C for 30 s followed by 6 min at 65°C (see Note 8).

3.2. Degradation of dNTPs and PCR Primers

Unused dNTPs and PCR primers from the PCR must be removed prior to the SBE reaction. In the method described below, exonuclease I (ExoI) degrades single-stranded DNA molecules, e.g., PCR primers, and SAP dephosphorylates dNTPs to dNMPs (see Note 9).

1. Pipette 5-µL amplified DNA from Subheading 3.1 to a new PCR plate. The remaining 20-µL amplified DNA may be stored at −20°C.

2. Add 2-µL ExoI-SAP solution.

3. Seal the PCR plate with adhesive seal, shake it briefly on a Vortex mixer and spin it briefly in a plate centrifuge at 600 rpm.

4. Place the PCR plate in a PCR thermocycler and run the following programme: 37°C for 60 min followed by 75°C for 15 min.

3.3. Multiplex SBE

1. Thaw and mix the reagents on a Vortex mixer before placing them on ice.

2. Make a reaction mixture A with 4-µL ABI Prism SNaPshot™ Multiplex Ready Reaction Mix, 2-µL nuclease-free water and 1-µL SBE primer mix A (Table 2) for each sample. The reaction mixture should be shaken on a Vortex mixer and placed on ice.

3. Make a reaction mixture B with 4-µL ABI Prism SNaPshot™ Multiplex Ready Reaction Mix, 2-µL ddH$_2$O and 1-µL SBE primer mix B (Table 3) for each sample. The reaction mixture should be shaken on a Vortex mixer and placed on ice.

4. Mix 7-µL reaction mixture A and 1-µL amplified DNA from Subheading 3.2 in a new PCR plate.

5. Mix 7-µL reaction mixture B and 1-µL amplified DNA from Subheading 3.2 in a new PCR plate. The remaining 5-µL amplified DNA from Subheading 3.2 may be stored at −20°C.

6. Seal the PCR plates with adhesive seal, shake them briefly on a Vortex mixer and spin them briefly in a plate centrifuge at 600 rpm.

7. Place the PCR plates in a PCR thermocycler (see Note 1) and run the following cycle programme: 30 cycles of 96°C for 10 s, 55°C for 5 s and 60°C for 30 s.

3.4. Degradation of ddNTPs

Unused ddNTPs from the SBE reaction gives a lot of signals in the electropherogram. In the method described below, the ddNTPs are degraded by SAP.

1. Add 1 µL 1 U/µL SAP to each well in the PCR plates from Subheading 3.3.

2. Seal the PCR plates with adhesive seal, shake them briefly on a Vortex mixer and spin them briefly in a plate centrifuge at 600 rpm.

3. Place the PCR plates in a PCR thermocycler and run the following cycle programme: 37°C for 30 min followed by 75°C for 15 min.

3.5. Capillary Electrophoresis

A spectral calibration using the DS-02 Matrix Standard (dye set E5) must be approved prior to electrophoresis of the SBE products.

1. Make a size standard mixture of Hi-Di™ formamide and GeneScan™ – 120 LIZ™ size standard in a ratio of 100:1.

2. Pipette 20-µL size standard mixture into each well of the electrophoresis plate.

3. Add 1-µL SBE products from Subheading 3.4.

4. Prepare the plate for electrophoresis and spin it briefly in a plate centrifuge at 600 rpm.

5. Denature the samples at 95°C for 3 min and cool them immediately by placing the plate in a container cooled down to 4°C.

6. Run the electrophoresis with the following settings (see Note 10).

Oven temperature:	60°C
PreRun voltage:	15 kV
PreRun time:	180 s
Injection voltage:	3 kV
Injection time:	22 s
Run voltage:	15 kV
Run time:	1,000 s

3.6. Analysis of the SBE Products

Analysis should be performed independently by two people and their results compared (see step 7 below).

1. Import the data files into GeneScan® 3.7.1 (see Note 11) and analyse the results with the following analysis parameters:

Size call range:	Full range
Smooth option:	None
Peak detection:	50 RFU for all colours
Minimum peak half width:	2
Polynomial degree:	3
Peak window size:	15
Slope threshold for peak start:	0
Slope threshold for peak end:	0
Size calling method:	Local Southern
Baselining window size:	51

2. Save the GeneScan analysis file and close GeneScan® 3.7.1.

3. Import the GeneScan analysis file into Genotyper™ 3.7 and label all relevant peaks with allele name, length and peak height (see Note 12).

4. Generate a text file with sample ID, allele calls and peak heights.

5. Open the text file in Microsoft Excel® and test the quality of the allele calls against the expected peak height ratios for heterozygote allele calls and the expected signal/noise ratio for homozygous allele calls (see Note 13).

6. Re-analyse the allele calls that were not approved in step 5 in Genotyper™ 3.7 and decide whether the allele call is correct or if the allele call should be failed (see Notes 14 and 15).

7. Compare the results of two independent analyses. If the results differ, the results should be re-analysed independently by both analysts and compared again. If the analysts disagree after the second round of analyses, the allele calls should be failed.

4. Notes

1. Alternative DNA polymerases, buffer systems and thermocyclers may be used. However, this will most likely require further optimization and balancing of the PCR and SBE reactions ((13), unpublished results). We have successfully used the GeneAmp® PCR system AB9600, the GeneAmp® PCR system AB9700,

the AB2720 thermal cycler, and the Eppendorf Mastercycler® gradient thermocycler.

2. Repeated thawing and freezing of dNTPs and primer mixes will reduce the performance of the assay. dNTPs and primer mixes should be stored in small aliquots suitable for one or two experiments.

3. High-quality oligonucleotides are essential for all multiplex reactions (13). In our experience, there is a large difference in primer quality between vendors and it is important to choose the right one. Most vendors perform quality analyses of their oligonucleotides by MALDI-TOF mass spectrometry. It is usually possible to get access to the mass spectra and they can be very helpful in a troubleshooting process or for selection of the best vendor.

4. The SNP*for*ID multiplex has been used successfully on genomic DNA purified by many different extraction protocols, including Chelex resin, phenol/chloroform, Proteinase K/Laurylsarcosin/EDTA/QIAquick PCR purification kit, QIAsymphony DNA Investigator kit, QIAamp DNA Investigator kit, EZ1 DNA Investigator kit, MagAttract DNA mini M48 kit and more. The SNP*for*ID multiplex has also been used successfully on buccal swabs or blood placed on FTA-cards.

5. The ExoI-SAP solution may be stored at +4°C for 5 days. However, it may increase the background noise from PCR primers extended with labelled ddNTPs in the SBE reaction and we recommend making a fresh solution for each experiment. The ExoI-SAP solution may be bought as a mixture, but that is more expensive than buying the enzymes separately.

6. POP-6™ and POP-7™ may also be used. Most forensic genetic applications use POP-4™ and it is convenient to use the same polymer for all assays performed in the laboratory.

7. PCR is a highly efficient method for amplification of DNA. However, the extreme sensitivity of PCR also makes the technique susceptible to contamination problems (14, 15). Laboratories performing PCR routinely need laboratory protocols to avoid and check for cross-contamination (16–18). These protocols are included in the accredited laboratory standards for testing laboratories (ISO 17025) recommended by the International Society of Forensic Genetics (ISFG) (19). It is essential to physically separate pre-PCR work areas from post-PCR work areas and to ensure that all traffic of reagents and, if possible, laboratory workers is unidirectional from the pre-PCR work area to the post-PCR work area. Each designated work area must be fully equipped with the necessary instruments, laminar air flow benches, laboratory disposables, laboratory coats, etc. Good laboratory behaviour must be

conducted in all areas. Thus, work stations must be cleaned with 10% freshly prepared bleach before and after each laboratory step, reusable trays and racks must be immersed in 10% bleach before they are cleaned and reused, the laboratory workers must wear laboratory coats and gloves at all times, and used disposables and reagents must be removed from the designated areas in closed containers. Furthermore, it is recommended to build pre-PCR work areas with positive air pressure and post-PCR work areas with negative air pressure.

8. In the multiplex PCR described here, the Mg^{2+} concentration is unusually high and the elongation temperature is unusually low compared to standard singleplex PCR conditions. This is done to optimize and balance the multiplex PCR. There are other ways to improve amplification efficiency in multiplex PCR, including the concentrations of the buffer, dNTPs, and PCR primers, primer annealing temperature and PCR primer design (13, 20–22).

9. Unused dNTPs and PCR primers may also be removed by various PCR purification kits. This is advantageous if the PCR products are going to be used in down-stream applications that are more sensitive, such as mass spectrometry (23), or if the down-stream application employs enzymes that are inactive in the buffer systems used for the multiplex PCR. However, with some of the PCR purification kits, the pore size of the filter is too big and the smallest PCR products may be lost. We have successfully purified fragments down to a size of 44 bp with the standard protocol for the MinElute™ PCR purification kit (24).

10. It is possible to inject the products of two (or more) SBE reactions sequentially without injection of polymer in between and without any carry-over from the previous injection to the next one. This allows the collection of all results from one sample in only one data collection file (1, 2). The run time of the SNP36_POP4_DefaultModule must be changed to 300 s to reduce the time between the two injections. For the second injection, another module must be created, where three washing steps are added and where the capillary polymer fill stage and pre-run is removed. The run time of the second injection must be 1,000 s.

11. GeneMapper® software may also be used for the analysis.

12. Allele windows for all possible alleles should be pre-defined in the Genotyper™ 3.7 template. These windows are usually relatively narrow (0.3–0.6 nucleotides) compared to allele windows for PCR products (e.g. STR alleles), because the extended SBE primers have the same basepair composition in every experiment (2). Allele windows for ddNTP-extended PCR products

should also be pre-defined to simplify the analysis process ([2]). Also, define windows that are 2–3 nucleotides shorter and longer than the smallest and longest allele window in each colour, respectively. This will ensure that all allele calls will appear in the text file. Allele calls that are shorter or longer than the defined allele windows will not appear in the text file. Check the positive and the negative control before the rest of the samples are analysed. The SNP profile of the positive control should be identical to results from previous experiments. There will be peaks in the negative control. They originate from SBE of primer-dimers and the peaks will not necessarily be the same from experiment to experiment. However, the negative control should not look like a sample profile and the few peaks that may be detected in the electropherogram should be low.

13. Results from inter-laboratory exercises conducted by the SNP*for*ID consortium ([25]) and the European DNA Profiling Group ([26]) revealed that simple misinterpretation of the electropherograms was the most common mistake made by the participating laboratories. Small-sized peaks from PCR products or PCR primers extended with ddNTPs in the SBE reaction were called as true alleles and small-sized peaks were accepted as homozygous allele calls even though there was a high risk of allele drop-out. These problems were addressed when the SNP*for*ID multiplex was validated for forensic case work ([2]) and a protocol for analysis of the electropherograms was described that include a quality check of each allele call based on the peak height of that allele. A set of guidelines for allele calling was developed from the duplicate typing of 216 individuals from Danish paternity cases and acceptable peak height ratios for heterozygote allele calls and acceptable signal/noise ratio for homozygous allele calls were defined ([2]).

14. There are many reasons why an allele call may be acceptable even though it does not fulfil the guidelines for allele calling (see Note 13). For homozygous allele calls, the noise in the relevant part of the electropherogram may be lower than the average maximum noise level defined previously ([2]) and the true signal/noise ratio may therefore be acceptable. For heterozygous allele calls, the noise in one colour may be higher than the noise in the other colour or the signal from one allele may be saturated. This may lead to peak height ratios that were not acceptable even though the allele calls were correct. However, it is important to be cautious at this step in the protocol and experience and training are required.

15. If the maximum peak height was less than 3,500 RFU or there were many unacceptable homozygous allele calls, it indicates that the PCR amplification was weak and it may be prudent to re-type the sample with more DNA input in the PCR.

For weakly amplified samples, there will also be many unacceptable heterozygous allele calls, because the RFU scale is sigmoidal. If there are many unacceptable heterozygous allele calls, it indicates that the sample is a mixture or that the sample is contaminated (2). We have observed examples of unusual peak height ratios that are reproducible and in some cases inheritable (2, 5, 7). Some of these were identified as partial null alleles, where one allele was weakly amplified because of a mutation in the PCR primer binding site. However, the partial null alleles are rare, and in mixtures, there are ten or more unusual peak height ratios which clearly indicate that the sample contains DNA from more than one individual.

References

1. Sanchez JJ, Phillips C, Børsting C et al (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27:1713–24

2. Børsting C, Rockenbauer E, Morling N (2009) Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard. Forensic Sci Int Genet 4:34–42

3. Rockenbauer E, Børsting C, Stangegaard M et al (2009) Successful STR and SNP typing of FTA-card samples with low amounts of DNA after DNA extraction using a Qiagen BioRobot® EZ1. Forensic Sci Int Genet Suppl 2:83–84

4. Fondevilla M, Phillips C, Naveran N et al (2008) Case report: Identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. Forensic Sci Int Genet 2:212–218

5. Børsting C, Sanchez JJ, Hansen HE et al (2008) Performance of the SNP*for*ID 52 SNP-plex assay in paternity testing. Forensic Sci Int Genet 2:292–300

6. Porras L, Phillips C, Fondevilla M et al (2009) Genetic variability of the SNP*for*ID 52-plex identification SNP panel in Central West Columbia. Forensic Sci Int Genet 4:e9-e10

7. Drobnic K, Børsting C, Rockenbauer E et al (2010) Typing of 49 autosomal SNPs by SNaPshot® in the Slovenian population. Forensic Sci Int Genet 4:e125–e127

8. Santos C, Phillips C, Fondevilla M et al (2010) A study of East Timor variability using the SNP*for*ID 52-plex SNP panel. Forensic Sci Int Genet, in press

9. Poulsen L, Børsting C, Tomas C et al (2010) Typing of Amerindian Kichwas and Mestizos from Ecuador with the SNP*for*ID multiplex. Forensic Sci Int Genet, submitted for publication

10. Amigo J, Phillips C, Lareu M et al (2008) The SNP*for*ID browser: an online tool for query and display of frequency data from the SNP*for*ID project. Int J Legal Med 122:435–440

11. Hansen AJ, Simonsen BT, Børsting C et al (2006) Semi-automatic preparation of biological database samples for STR and SNP typing. Progr Forensic Genet 11:663–665

12. Stangegaard M, Olsen AN, Frøslev TG et al (2009) Automated washing of FTA card punches and PCR setup for reference samples using a LIMS-controlled Sias Xantus automated liquid handler. Forensic Sci Int Genet Suppl 2:69–70

13. Sanchez JJ, Børsting C, Morling N (2005) Typing of Y chromosome SNPs with multiplex PCR methods. In: Carracedo A (ed) Forensic DNA typing protocols, Meth Mol Biol 297:209–228, Humana press, New Jersey

14. Persing DH (1991) Polymerase chain reaction: Trenches to benches. J Clin Microbiol 29: 1281–1285

15. Borst A, Box ATA, Fluit AC (2004) False-positive results and contamination in nucleic acid amplification assays: Suggestions for a prevent and destroy strategy. Eur J Clin Microbiol Infect Dis 23:289–299

16. Kwok S, Higuchi R (1989) Avoiding false positives with PCR. Nature 339:237–238

17. Victor T, Jordaan A, du Toit R & Van Helden PD (1993) Laboratory experience and guidelines for avoiding false positive polymerase

chain reaction results. Eur. J. Clin. Chem. Clin. Biochem. 31:531–535

18. Burkardt H (2000) Standardization and quality control of PCR analyses. Clin Chem Lab Med 38:87–91

19. Morling N, Allen RW, Carrecedo A et al (2002) Paternity testing commission of the International Society of Forensic Genetics: recommendation on genetic investigation in paternity cases. Forensic Sci Int 129:148–157

20. Henegariu O, Heerema NA, Dlouhy SR et al (1997) Multiplex PCR: Critical parameters and step-by-step protocol. Biotechniques 23: 504–511

21. Markoulatos P Siafakas N, Moncany M (2002) Multiplex polymerase chain reaction: A practical approach. J. Clin Lab Anal 16:47–51

22. Multiple authors (2007) Multiplex PCR primer design. In: Anton Yuryev (ed) PCR primer design, Meth Mol Biol 402:Chapter 11–15, Humana press, New Jersey

23. Mengel-Jørgensen J, Sanchez JJ, Børsting C et al (2005) Typing of multiple SNPs using ribonuclease cleavage of DNA/RNA chimeric single base extension primers and detection by MALDI-TOF mass spectrometry. Anal Chem77:5229–5235

24. Børsting C, Sanchez JJ, Morling N (2004) Multiplex PCR, amplicon size and hybridization efficiency on the NanoChip electronic microarray Int J Legal Med 118:75–82

25. Musgrave-Brown E, Ballard D, Balogh K et al (2007) Forensic validation of the SNP*for*ID 52-plex assay. Forensic Sci Int Genet 1: 186–190

26. Sanchez JJ, Børsting C, Balogh K et al (2008) Forensic typing of autosomal SNPs with a 29 SNP-multiplex – results of a collaborative EDNAP exercise. Forensic Sci Int Genet 2:176–183

Chapter 8

A 34-plex Autosomal SNP Single Base Extension Assay for Ancestry Investigations

C. Phillips, M. Fondevila, and Maria Victoria Lareau

Abstract

Ancestry inference based on autosomal markers remains a niche approach in forensic analysis: most laboratories feel more secure with a review of the cumulative STR profile frequencies in a range of relevant populations with the possible additional analysis of mitochondrial and/or Y-chromosome variability. However, a proportion of autosomal single nucleotide polymorphisms (SNPs) show very well-differentiated allele frequencies among global population-groups. Furthermore, such ancestry informative marker SNPs (AIM-SNPs) lend themselves to relatively straightforward typing with short-amplicon PCR and multi-plexed single base extension reactions using the same capillary electrophoresis detectors required for the sequencing and STR genotyping of mainstream forensic markers. In this chapter, we describe a 34 AIM-SNP multiplex that is robust enough for the analysis of challenging, often highly degraded DNA typical of much of routine forensic casework. We also outline in detail the in-silico procedures necessary for collecting parental population reference data from the SPSmart SNP databases and performing ancestry inference of single AIM-SNP profiles or large-scale population data using the companion ancestry analysis website of *Snipper*. Two casework examples are described that show, in both cases, that an inference of likely ancestry using AIM-SNPs helped the identification of highly degraded skeletal material.

Key words: Genetic ancestry, SNP, AIM-SNP, Population genetics, Population admixture, Bayesian analysis

1. Introduction

Although typing of ancestry informative marker-single nucleotide polymorphisms (AIM-SNPs) is widely used in the study of ancestry for medical and population genetics, their application to forensic analysis has been much more tentative. This is due in part to the fact that the AIM-SNPs typed must be components of relatively small-scale multiplexes, ranging from 20 to 35 markers, to be able

to work efficiently with restricted levels of DNA. Such limited AIM-SNP numbers mean the most informative markers must be selected and combined to maximize the power to differentiate populations. In addition, methods of analysis that permit a reliable inference of ancestry from the observed patterns of allelic variation are often perceived to have less statistical value compared to the levels of profile matching probabilities routinely generated for individual identification. Lastly, population admixture, a significant characteristic of most urban populations of the Western World, acts to erode the differentiations observed in the unadmixed parental populations used as reference data to make inferences of individual ancestry for a SNP profile. In this context, parental population data signify genotypes found in component populations of the five global population-groups of Africa, East Asia, Europe, The Americas, and Oceania (herein AFR, E ASN, EUR, AME, OCE). These five groups largely correspond to the geography of the major continental regions and they have been shown to be genetically distinct by an extensive body of work, of which Rosenberg et al. (1) arguably represents the landmark publication of such studies. In fact the five major groups are bound more by geographic barriers than continental margins, for example North African populations show a stronger affinity to European populations than those of sub-Saharan Africa, south of the Sahel. Similarly, South Asian populations are more closely related to Europeans than East Asians in the absence of significant geographic barriers in central Asia. Therefore, the meta-population covering regions of Europe across to Central-South Asia (comprising Europeans, Middle East-North Africans, South Asians) can be more appropriately referred to as Eurasian.

At the time of writing, three autosomal AIM-SNP multiplexes specifically designed for the analysis of forensic DNA have been described in the literature (2–4). All three publications list the primer sequences along with multiplex details, enabling users to assess alternative approaches for simple ancestry analysis of forensic DNA. The methods of Phillips et al. and Lao et al. (2, 3) use SNaPshot assays, while Halder et al. (4) use *SNPstream* technology, but all show forensic levels of sensitivity while the 34-plex assay of Phillips et al., described in this chapter, has been observed to show particular sensitivity with challenging DNA (5, 6). Furthermore, the *in-silico* analytical method of the 34-plex assay includes a simple statistical system available at the *Snipper* web portal [http://mathgene.usc.es/snipper/] to assign a likely ancestry to the SNP profile generated, but more importantly, this is now extendible to any SNP set for which comparative parental population data exist.

The most popular approach for ancestry analysis of genotype data from multiple populations is the *structure* algorithm (1, 7)

(see Note 1). In particular, *structure* is the method of choice for the analysis of admixture since one of its strengths lies in the ability to accurately gauge the components of admixture per individual, especially when the contributor populations are well differentiated by the markers typed. In addition, *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. However, *structure* analysis of single individual profiles is not so straightforward since the whole set of parental data plus the unknown profile must be 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* was developed to provide a simple alternative for making ancestry assignments of single profiles in real-time, rather than the more computationally intensive and slower process of assessing the whole data set each time with *structure*. Both *structure* and *Snipper* use a naïve Bayesian analysis which, put simply, equates likelihood of membership of a class (in this case ancestry) directly to the observed frequency of variables in each class (in this case allele frequencies). Therefore, both algorithms require reference data from which to calculate the allele frequencies and then compare these to the alleles found in the unknown profile(s). In the case of *Snipper,* the reference data allow construction of training sets from which to calculate allele frequencies and these sets can comprise fixed three group data (African, East Asian, European) in place for 34 SNPs and ready to use or can be data for any populations and SNP 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 (hence naïve): that the variables, i.e., the component loci, are independent. So uni-parental data are not readily incorporated into either analysis system, though *structure* has scope for the analysis of linked loci.

In the following sections, we describe: (1) how to obtain parental population reference data for AIM-SNPs to use with either *structure* or *Snipper*; (2) how to use *Snipper* to analyze a 34-plex AIM-SNP profile using a simple fixed three group training set; (3) how this approach can be extended to analyze custom data with an example of the 34 SNPs of the standard forensic assay supplemented with an additional 24 SNP multiplex designed to differentiate South Asians from Europeans and; (4) how to quickly assess the ancestry of multiple profiles. In conclusion, we briefly review the potential in the future of our advocated approach to forensic ancestry analysis: mixed marker tests that combine STR and AIM-SNP genotypes from the same forensic case to make the most robust inference of ancestry from the available DNA data.

2. Materials

1. *Snipper* portal: http://mathgene.usc.es/snipper/default3_34_new.html

2. *Structure* software can be downloaded at: http://pritch.bsd.uchicago.edu/structure.html

3. SPSmart home: http://spsmart.cesga.es/

4. SPSmart SNP*for*ID 52-plex and 34-plex variability browser: http://spsmart.cesga.es/snpforid.php

5. Entire genome interface for exploring SNPs (ENGINES) a 1,000 genomes variability browser enabling a review of all SNP sites found from 1093 (more due in 2012) complete genome sequences: http://spsmart.cesga.es/search.php?dataSet=engine

6. pop.STR: http://spsmart.cesga.es/popstr.php

3. Methods

3.1. The 34-plex AIM-SNP SNaPshot Assay

The chemistry of the SNaPshot single base extension assay is well covered in other chapters and this section therefore just briefly covers additional observations from use of the 34-plex genotyping assay in routine forensic casework found since original publication (2). At the time of writing, the NIST 2391b standard reference material (https://www-s.nist.gov/srmors/view_detail.cfm?srm=2391B) is being genotyped and data will be made available to potential users by publication and in the SPSmart SNP*for*ID pages.

1. The 34-plex AIM-SNP assay uses the same primers as those listed in Phillips et al. (2). For guidance, Table 1 lists the observed average sizes in basepairs (bp) for the composite SNPs using POP-6 with a 3130 detector. Table 2 gives the ratios by volume of the components of the PCR and extension multiplex primer mixes. Users may need to adjust these depending on the observed peak height ratios following experimental runs of new primer stock combinations.

2. Standardized rules for the interpretation of heterozygote peak imbalances observed using SNaPshot dyes and previously established for the 52plex ID-SNP multiplex by The SNP*for*ID Consortium (8). Briefly these are: the blue G signal is 4 times greater or less than the yellow/red C/T signal and 2 times greater or less than the green A signal. There are two outlier heterozygote patterns in the 34-plex that often show greater imbalances than the above rules of interpretation: SNPs rs2304925

Table 1
Observed basepair sizes of 34-plex component SNPs at the developing laboratory (2), using an AB 3130 genetic analyzer and POP-4 or POP-6 polymer in a 50-cm capillary array

POP-4				SNP (mobility order)	POP-6			
G	A	C	T		G	A	C	T
22.95			25.76	rs2304925	19.69			21.37
	24.63	23.44		rs5997008		21.57	20.38	
		26.62	28.48	rs1321333			23.75	24.92
		29.45	30.26	rs2814778			24.56	26.02
26.25	28.78			rs917118	24.58	26.19		
28.16	30.11			rs1024116	26.7	28.19		
		31.34	32.62	rs7897550			28.74	29.85
33.86	36.08			rs722098	32.51	34.36		
		37.07	38.36	rs10843344			33.1	34.17
		38.9	40.4	rs239031			37.05	37.96
40.11	40.71			rs12913832	38.42	39.09		
43.06	44.08			rs1978806	41.33	42.22		
		44.18	45.41	rs2040411			42.17	43.1
46.37		46.76		rs773658	44.18		45.08	
	48.69		49.44	rs10141763		47.05		47.51
		49.21	50.52	rs182549			48.23	49.53
50.29	51.19			rs1573020	48.94	49.7		
		53.13	54.1	rs896788			52.01	52.88
54.9	55.78			rs2065160	53.46	54.27		
56.43	57.46			rs2572307	55.59	56.36		
		58.59	59.25	rs2303798			57.25	58.03
61.07	61.81			rs2065982	60.15	60.93		
		62.27	63.44	rs3785181			60.95	61.89
63.61			65.49	rs881929	63.24			64.52
	65.8	65.62		rs1498444		65.07	64.9	
69.72	70.45			rs2026721	68.79	69.47		
		68.9	69.43	rs1426654			68.18	68.85
	72.91	72.88	73.28	rs4540055		72.31	71.91	72.35

(continued)

Table 1
(continued)

	POP-4			SNP (mobility order)		POP-6		
G	A	C	T		G	A	C	T
	77.96		78.38	rs1335873		75.47		76.04
78.32		78.36		rs16891982	77.22		77.47	
		79.98	80.34	rs1886510			79.16	79.49
79.92	80.28			rs730570	79.55	80.25		
83.26	84.41	84		rs5030240	83.11	84.4	83.6	
	88.3	87.55		rs727811		87.66	86.87	

Table 2
PCR and extension primer ratios for the preparation of multiplex reaction components

	PCR		EXT	
	2 × 25 µM primer stock µL	Final concentration (µM)	25 µM primer stock µl	Final concentration (µM)
rs2304925	4	0.9425	2.5	1.1489
rs7897550	2	0.4713	1	0.4596
rs773658	3.2	0.7540	3	1.3787
rs239031	0.5	0.1178	1	0.4596
rs5997008	1.9	0.4477	1.3	0.5974
rs2040411	1.8	0.4241	1	0.4596
rs10141763	3	0.7069	1.15	0.5285
rs2572307	1.5	0.3534	1.1	0.5055
rs2026721	3.2	0.7540	1.75	0.8042
rs2303798	4.3	1.0132	3.5	1.6085
rs1573020	0.5	0.1178	1.25	0.5744
rs2065160	0.8	0.1885	1.3	0.5974
rs896788	1.5	0.3534	1.25	0.5744
rs2065982	2	0.4713	1.35	0.6204
rs10843344	3	0.7069	1.25	0.5744
rs1498444	5	1.1781	3	1.3787

(continued)

Table 2
(continued)

	PCR		EXT	
	2 × 25 μM primer stock μL	Final concentration (μM)	25 μM primer stock μl	Final concentration (μM)
rs4540055	1.2	0.2828	1.1	0.5055
rs182549	6.2	1.4609	6.5	2.9871
rs3785181	1.25	0.2945	0.75	0.3447
rs2814778	1.2	0.2828	2	0.9191
rs1335873	0.75	0.1767	1.6	0.7353
rs1886510	1.5	0.3534	1	0.4596
rs917118	1	0.2356	2	0.9191
rs722098	2.3	0.5419	4	1.8382
rs1978806	1	0.2356	2	0.9191
rs16891982	1.2	0.2828	1.75	0.8042
rs12913832	3	0.7069	3	1.3787
rs1321333	2.75	0.6480	1.5	0.6893
rs5030240	2	0.4713	2	0.9191
rs727811	2	0.4713	2.5	1.1489
rs1024116	3	0.7069	2.1	0.9651
rs730570	1.3	0.3063	2.2	1.0110
rs1426654	3	0.7069	4	1.8382
rs881929	3	0.7069	2.5	1.1489
H_2O	109.6		H_2O	6.27
			NH_3SO_4 (200 mM)	33.33
Total	179.45		Total	108.8

(red T less than 1/4 of the blue G signal strength) and rs12913832 (very strong blue:green imbalance likely due to a long-standing observation of poor extension of the A allele for this SNP).

3. Three persistent nonspecific peaks occur in the green channel at fixed positions 64.33, 82.19, and 87.14 bp (POP-6). These peaks are close to SNPs rs1498444, rs730570, and rs727811 with the 64.33 bp peak closer than a single base interval to the A allele of rs1498444, but all are easily distinguishable as the nonspecific signals always run to the same position and are generally lower than extension product peak heights.

3.2. Obtaining Reference Population AIM-SNP Data from SPSmart

A statistical analysis of a 34-plex profile with fixed AFR-EUR-E ASN training sets can be made immediately the profile genotypes are uploaded, but other population analyses require the user to collect and input reference data for the parental populations to be compared. This process previously required locus-by-locus scrutiny of dbSNP or HapMap SNP databases (9), but luckily SPSmart (10, 11) now makes the task much more straightforward for any number of AIM-SNPs. The SPSmart population-variation browser suite currently includes four components: a SNP browser for the SNP*for*ID forensic SNP sets (52-plex ID-SNPs and 34-plex AIM-SNPs); a browser for HapMap, Perlegen, and Stanford/Michigan CEPH-HGDP (see Note 2) SNP repositories; a browser for the 1,000 genomes SNP repository and; an STR browser initially based on the 20 core forensic markers and the 51 CEPH-HGDP populations. Note 3 lists the four URLs for direct access to the search pages of each browser.

1. Choose the database(s) to search or opt for SEARCH in the SNP*for*ID "global map" homepage. If opting to review multiple databases, e.g., HapMap and Perlegen, only one population grouping can be made.

2. Choose the populations to merge into groups by ticking selections—the maximum number of separate groups from a single database is five, but the same process can be repeated multiple times if more groupings are required. Populations are already arranged into sets of genetic diversity based on Rosenberg's original structure analyses of the CEPH-HGDP populations (1, 12), though it should be noted that the Eurasian meta-population has been subdivided into three subgroups.

3. Add the SNP RefSeq (rs-number) identifiers in the search by SNP window. There is no maximum number, but long lists can slow the query. Click the "next" button below.

4. Select or ignore the filter options presented. MAF is minimum allele frequency and *In* is Rosenberg's AIM informativeness metric (13). When reviewing SNP data from multiple databases, it is best not to tick option: "Filter SNPs not genotyped on every compared dataset."

5. In the downloads tab are the genotype files ready to download, then copy and paste into Excel or notepad. The recommended steps being: download, choose all, copy into Excel, invert the data into rows = samples and columns = SNPs (copy > select new cell > paste special > 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.

6. The genotypes can be formatted for input to *structure* or *Snipper* custom data analysis (outlined in previous paragraph).

SNaPshot genotypes may need checking against the reference data if each represents 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 4).

3.3. Making an Ancestry Inference from a 34-plex SNP Profile—A Simple Three Population-Group Comparison

The *Snipper* portal was originally designed to provide a real-time ancestry assignment system for 34-plex profiles with reference to default pretyped AFR-EUR-E ASN training sets and this still represents the simplest approach for assessment of a single casework profile requiring an immediate overview of ancestry. The portal can be accessed directly (http://mathgene.usc.es/snipper/default3_34_new.html) or via the homepage by selecting the link: "Classification as Europ.-Afr.-E. Asian (34 SNPs)." An assignment is made on the basis of three likelihood ratios (AFR:EUR, AFR:E ASN, EUR:E ASN) using three ancestry probabilities derived from the reference data (AFR: Mozambican and Somali data, EUR: Danish and NW Spanish or Galician, E ASN: Taiwanese and Chinese, 60 each population). Figure 1 describes a casework example

TTCCCCTTGGAATTCTNNTTAAAATTCCAACTAACCAAGGCCAGCCGTAATTGGATAACGTTAACCNN

Data source used for training sets in Snipper		Output from Snipper			
		-log lk	EXP values	likelihood ratios from EXP values	Verbal summary
Fixed 3-group training set of AFR-EUR-E ASN data	European	25.57249	7.83E-12	18,042,414,120,482,400,000	18 quintillion times more likely EUR than AFR
	African	69.91174	4.34E-31	3,137,090,832,720	3 trillion times more likely EUR than ASN
	East Asian	54.3468	2.50E-24		

Fig. 1. Casework example 1: Simple 34-plex analysis of challenging material. DNA extracted from burnt muscle (1) and molar dentine (2) taken from badly charred cadaver (3, 4) found in burnt flat. Police investigation sought to exclude the possibility of an itinerant from Africa, common in the area. A 32/34 AIM-SNP profile, shown *top-left*, was obtained from the dentine extract and a comparison of the profile to fixed three-group training sets comprising Danish/NW Spain (EUR), Mozambican/Somali (AFR) and Chinese/Taiwanese (E ASN) indicated a person 18 quintillion and three trillion times more likely to be European than African or East Asian, respectively.

using the fixed training set option of *Snipper* to identify a badly charred corpse found in a flat in Spain where it was important to exclude the possibility that the body was an itinerant from Africa.

1. Choose the order of genotypes in the submitted profile: SNaPshot mobility order or rs-number order.

2. 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.

3. Each missing genotype is entered as two Ns per SNP, so only ACGTN characters are permitted. Format errors from incorrect bases (either due to incorrect SNP order or inverted bases) are flagged by the portal with a warning for the relevant SNP position(s), e.g., *Error: SNP A30 (number 34) has possible alleles AC but your profile has first allele G in this SNP (global position 67). Please curate your profile.*

4. Results returned from a submission comprise: the submitted profile; the -log likelihoods and percentiles for the three training set groups; the likelihood ratios in verbose format and predicted ancestry; the apparent success and; a list of the composite SNPs in descending order of three group divergence (these values are a variant of Rosenberg's *In*).

5. Missing SNP 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 SNPs at the top of the list will be much less reliable than those with gaps at the bottom—though this will be clear from the probabilities obtained.

6. Apparent success measures the rate of correct assignment of training set samples using the SNPs of the profile. These values are 100% for a complete set of 34 SNPs, but drop when significant numbers of gaps occur in the submitted profile – particularly for EUR:E ASN comparisons.

3.4. Making an Ancestry Inference from any SNP Profile: Specialized Analyses Using Custom SNP and Population Data

In order to extend ancestry analyses beyond the default three population-group comparisons and 34 SNPs using *Snipper*, select the homepage red links: (1) "Classification with a custom Excel file of populations" and (2) "Thorough analysis of population data with a custom Excel file." Respectively, these allow: (1) ancestry assignment of a profile comprising any SNPs (up to 100 markers) using training sets of the same loci for populations defined by the user based on collected data or generated from population samples relevant for the laboratory, and (2) analysis of the training sets to gauge characteristics of the component SNPs. The latter analysis option is a useful aid to assessing the informativeness of new candidate AIM-SNPs for ancestry inference. A key factor in following

CTTTNNCTCCNNNNNNCTNNGGTTGGAAGGTTAAAAAGAGCTGGNNAAGGNNAGCCAATTCCAAGGCCGGTTGGNNAGAACCGTTTGGTTCCACCCCCNNNNTTCGCTGTNNAA

Data sources used to construct training sets uploaded to *Snipper*	Two most relevant likelihood values for ancestry affiliation to the training sets used in analysis.
SPSmart CEPH-HGDP data for 7 groups (inc. ME and S ASN)	**Profile is 5.48x more likely MIDDLE EAST than EUROPE;** 9,524x more likely MIDDLE EAST than CS ASIA
CEPH 3 group (AFR-EUR-E ASN) + in house 2 group (ME-S ASN)	**Profile is 519x more likely MIDDLE EAST than EUROPE;** 3,776,408x more likely MIDDLE EAST than CS ASIA
CEPH 3 group (AFR-EUR-E ASN) + in house S ASN data	**Profile is 7,277x more likely EUROPE than SOUTH ASIA;** 526,949,673,156,779,119,214,592x more likely EUROPE than EAST ASIA

Fig. 2. Casework example 2: Custom analysis of 34-plex data supplemented with a custom AIM-SNP set of 24 additional markers. Skeletal remains (1, 2) were found washed ashore in La Coruña harbor (3) and DNA was extracted from a molar visible in right mandible. Biometrical analyses of skull dimensions suggested South Asian rather than European origin. A genotype profile of 47/58 AIM-SNPs was obtained and analyzed in *Snipper* comparing three parental datasets as custom training sets: (a) SPSmart CEPH-HGDP AFR, E ASN, AME, OCE plus the European, Middle Eastern (ME) and S Asian (S ASN) components of Eurasia; (b) CEPH AFR, E ASN, European plus in-house genotypes from Iranian/Iraqi (ME) and Indian (S ASN) samples, and (c) genotype data of ii. without ME samples. Ancestry likelihoods show the individual is much more likely to be European or Middle East in origin than South Asian. Although cross-validation revealed a zero misclassification with S ASN vs. ME or European, there was less success differentiating ME from European; therefore, this data was interpreted just to infer Eurasian ancestry outside of South Asia, directing the investigation towards missing European and Middle Eastern merchant seamen.

both the above analysis options is careful preparation of the Excel file containing the custom training set profiles and precise matching of unknown profiles to training set profiles in terms of bases and locus order—therefore it is recommended to sort component SNPs into ascending rs-number order as an aid to data checking.

Figure 2 describes a second casework example where standard 34-plex AIM-SNP analysis was supplemented with a custom SNP set of markers chosen in order to differentiate European and South Asian ancestry. This was relevant to the identification of scant skeletal remains washed ashore in northern Spain, where biometrical measurements of the skull found suggested a possible South Asian ancestry and it was necessary to check records of missing merchant seamen of various geographic origins.

1. Download the example file demonstrating the required format of the Excel file of custom data (see Note 5).

2. Following the example arrangement and format (e.g., only ACGTN characters, etc.), compose a file of reference population profiles for the SNPs of interest obtained from SPSmart queries or in-house genotyping. Any reasonable number of populations can be compiled as sets of rows and up to 100 SNPs can be compiled as columns.

3. Choose the second link: "Thorough analysis of population data with a custom Excel file." Upload the Excel file of custom data and 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 of "classification of all individuals in the sample," "non-verbose cross validation," and "bootstrapping" are alternative assignment error estimators that can also be chosen.

4. "Do calculations" returns a ranked divergence list and the individual ancestry likelihoods for each sample assigned to the populations defined in the custom file. The accompanying prediction is listed, and when incorrect, shows four stars on the right along with missing SNP data for that sample. The total success rate in each training set is summarized in a reclassification table at the base of the results page.

5. As a simple "walkthrough," try collecting AFR (choosing MKK and YRI only to avoid incomplete coverage), EUR, and E ASN HapMap data from SPSmart for markers rs1426654 and rs12075, then compile the 3-population, 2-SNP profiles into a custom file for Snipper. There should be 309 AFR, 223 EUR, 364 E ASN, and cells A1, B1, C1 should hold values: 896 (total training set samples/rows), 2 (SNPs), 3 (populations compared). Once submitted and analyzed, note that despite using just two AIM-SNPs the majority of samples in each training set are correctly assigned. Although AFRs have a low success rate of 65.37%, E ASNs are correctly assigned in almost 100% of cases—reflecting the near-fixed allele frequency distributions shown by the pie charts of SPSmart in the selected populations.

6. Once the custom training set has been assessed for effectiveness, users can choose "classification with a custom Excel file of populations" by uploading matching single profiles from unknown samples to compare to the custom reference data and assigning ancestry in a similar fashion to the 34-plex fixed training set approach.

7. In the casework example given in Fig. 2, the 34-plex data were compiled into a custom training set file from the SPSmart SNP*for*ID pages and combined with additional custom SNP data obtained from the SPSmart Stanford/Michigan browser for the same CEPH-HGDP populations. However, when population coverage is "patchy," i.e., some SNPs are not typed in all populations of a database, then data can be collated in complimentary geographic regions, e.g., HapMap Yoruba AFR (YRI) plus Perlegen European American EUR and Han Chinese E ASN when either database has incomplete population coverage.

3.5. Multiple Profile Ancestry Analysis with Snipper

When large numbers of AIM-SNP profiles have been prepared from samples of unknown origin, it is sometimes convenient to make a simple, rapid check of the pattern of ancestries in the set without initiating a potentially lengthy *structure* analysis. This can be achieved with *Snipper* by including the unknowns in a custom training set file and performing a cross-validation as outlined in Subheading 2, item 4. Appropriate reference data can be included each time—for example, to check for ancestry outliers among case and control samples in a South American study sample, it would be prudent to include European and Native American reference populations but also include a comparative analysis of African parental data.

1. Label the unknown profiles in the custom file with the same name as one of the reference populations and perform a cross-validation. In the results, identify any samples assigned to other populations—this procedure may need repeating with alternative population labels for the unknowns if the majority do not match their first assigned population.

2. Pairwise charts that compare two ancestries can be made to enable a simple comparison of the range of likelihoods observed in the unknowns alongside their closest parental population vs. alternative populations. The charts are made in Excel by converting the *Snipper* likelihoods to whole numbers (using the = EXP formula in Excel), then making each pairwise ratio (e.g., lk EUR/lk AFR) and ranking the values obtained in descending order. Charting these with a log 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 (a ratio value meaning balanced odds of ancestry assignment to either population). The most distant points from midline represent the strongest assignments. In populations without admixture, the points are fully separated; when admixture occurs, a significant proportion of values are close to or cross the midline.

3. Figure 3 shows an example plot based on a custom data set of 46 AIM-SNPs informative for Native American and European ancestry. The EUR:AME likelihood ratio values are plotted for 50 HapMap Europeans from Utah (CEU) and 50 Individuals with Mexican ancestry from Los Angeles (MEX). The broad range of MEX likelihoods, mainly in the EUR upper side, but close to the line and crossing into the AME lower side, suggests admixed ancestry between European and American contributing populations applies to most of the samples with just 10% showing comparable ancestry probabilities to CEU and 14% showing predominantly American ancestry. The rest are admixed with a dominant European contribution with fewer showing more equal contributions from both parental populations.
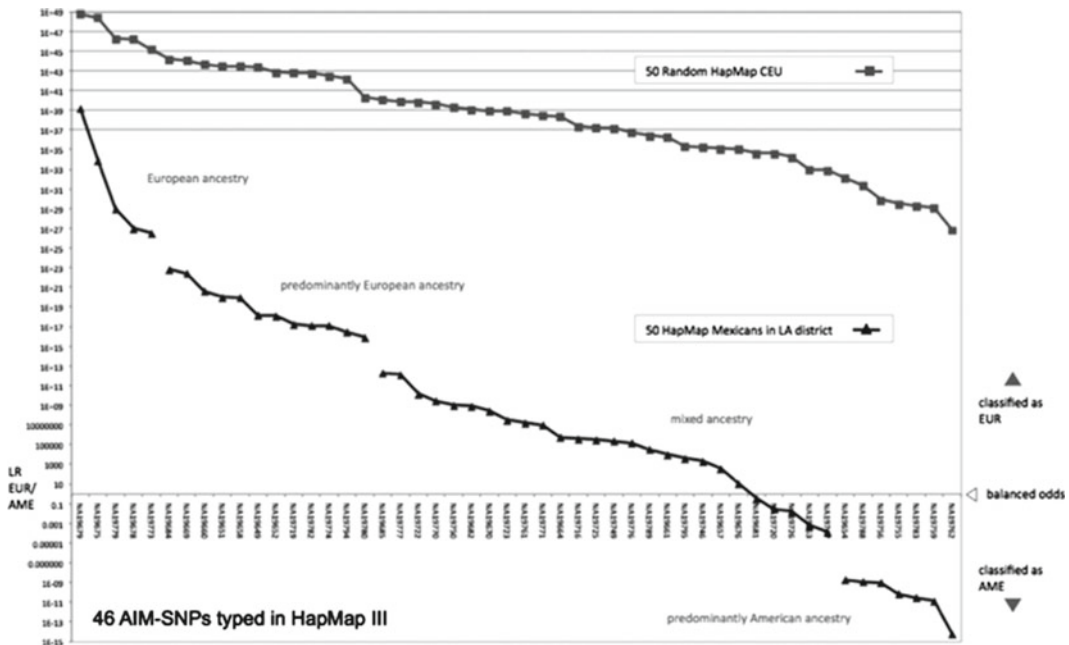
Fig. 3. Section of a pairwise plot showing EUR:AME likelihood ratio values from 46 AIM-SNPs from which HapMap data was collectable in SPSmart. Fifty HapMap Europeans from Utah (CEU) and 50 Individuals with Mexican ancestry from Los Angeles (MEX) are plotted, but the CEPH American data with which they were compared is not shown. The broad range of MEX likelihoods, noticeably lower in all but five cases than those of CEU Europeans, is strongly indicative of admixed ancestry between European and American contributing populations with the European component dominant in all but 10–12 samples, of which only seven in total have predominantly Native American ancestry.

Though a more detailed analysis can be achieved with *structure* and the reliability of interpretations is dependent on the level of differentiation shown by the AIM-SNPs used, such population snapshots are easily accomplished in real time using the custom data options of *Snipper*.

3.6. Concluding Remarks: Moving Towards Forensic Ancestry Analysis with Mixed Marker Approaches

Several studies indicate that STRs, locus for locus, are more ancestry informative than most SNPs (13–15). The term "most" here is an important qualification as clearly the best AIM-SNPs: those markers showing fixed or near-fixed allele frequency differences between populations will be the most informative loci of all and allele fixation in STRs is much rarer or the fixed alleles themselves are rare. In fact, D9S1120 (16) is the only mainstream STR characterized to date that exhibits a common allele exclusively found in one population-group—Native Americans. Perhaps, it is more important to consider the strategy of using all available DNA information in a forensic case as contributing factors towards building a more broadly based and thus more reliable prediction of the likely ancestry of a sample. STRs by themselves are not sufficiently informative to provide good differentiation of all five population-groups (17),

while a small set of AIM-SNPs by themselves will not be informative enough in every case—notably they can lack power in small numbers when extending comparisons to many diverse potential ancestries or to closely related populations (6). Therefore, it is logical to combine both marker sets if these have been characterized as part of the forensic analyses made in a case. The *structure* algorithm allows straightforward combinations of SNP genotype columns alongside STR genotype columns. Until now, *Snipper* has been focussed exclusively on SNPs, but with the launch of pop. STR, a browser listing STR data for the CEPH-HGDP populations dispersed across all geographic regions, the opportunity now exists to use this data to construct training sets comprising SNPs and STRs.

One problem is the paucity of STR genotype profiles for making training sets since published population studies report just allele frequencies, while release of STR genotypes suggests the possibility, however remote, of database searching, potentially compromising the anonymity of population study donors. For this reason, we have developed an allele frequency based input method for *Snipper* to allow users to upload their own frequencies, those obtained from sites such as SPSmart, taken from publications, or combinations of all three sources. STRs present the challenge of dealing with much more extensive allelic ranges—even in loci such as CSF1PO, we intend to designed custom data input files as Excel files with tabs for each marker holding columns for each allele and rows for each population (see Note 6). In this way, binary SNPs or indels can be included alongside as simplified two column tabs and uploaded as a single file of all markers typed. Evidence from *structure* analysis of the CEPH-HGDP (17) already suggests that the 34-plex AIM-SNP assay described in this chapter is enhanced considerably by combination with 15–20 standard identification STRs.

4. Notes

1. *Structure* software can be downloaded at: http://pritch.bsd. uchicago.edu/structure.html and additional information can be found in the help file (documentation for structure software: Version 2.3) included in the software folder. A forum devoted to *structure* is also running at: http://groups.google. com/group/structure-software?pli=1. Furthermore, the original article describing the algorithm Pritchard et al. (7) plus those outlining enhancements to the method: Falush et al. (18, 19); Hubisz et al. (20), are strongly recommended.

2. The CEPH-HGDP (Human Genome Diversity Panel) is a cell-line DNA set of 944 samples with wide currency in population genetics studies. The geographic coverage is patchy in certain

regions, but all continents and all major genetic ancestry groups defined by studies of variability are represented, the smallest numbers being 28 Oceanians from two populations and just six San from Namibia. There is freely accessible and downloadable genotype data from 650,000 SNPs for each CEPH-HGDP sample in the Stanford and Michigan University options within SPSmart. Both studies represent the most extensive coverage of population diversity so far achieved, though 1,000 genomes will extend to all sequence variation down to a likely minimum SNP allele frequency of 1% and are planned to go well beyond 1,000 samples from four continents (no OCE donors have been studied to date) during 2011–2012. The ENGINES browser in SPSmart (see Note 3) allows scrutiny of the first 178 whole sequences from African, European, and East Asian samples previously used to make the HapMap common-variant SNP catalog.

3. SPS (SNPs for population studies) comprises a group of variability browsers focussed on SNPs, but now also including forensic STRs, with four main components:

   - SPSmart home: http://spsmart.cesga.es/.
   - SPSmart SNP*for*ID 52-plex and 34-plex variability browser: http://spsmart.cesga.es/snpforid.php.
   - ENGINES (Entire Genome Interface for Exploring SNPs) a 1,000 genomes variability browser, enabling a review of all SNP sites found from 178 (more due in 2011) complete genome sequences: http://spsmart.cesga.es/search.php?dataSet=engine.
   - Pop.STR: http://spsmart.cesga.es/popstr.php.

4. 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 (or 52-plex) EXT 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 as shown by 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.

5. The custom data input file must be in .xls format, not the more recent xml format of .xlxs. Rows 2–5 can be left blank, row 1 must contain total sample (cell A1), total SNPs (A2), and total populations compared (A3). Cells A4 onwards contain the rs-number or users alternative SNP descriptors. It is sometimes necessary to "clean" the file by removing blank rows at the bottom of the data block and blank columns to the right—just highlight and delete 10–20 of each.

6. Classification with Snipper using frequencies alone is available at: http://mathgene.usc.es/snipper/frequencies.html.

References

1. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. (2002) Genetic structure of human populations. Science 298: 2381–5.

2. Phillips C, Salas A, Sánchez JJ, Fondevila M, Gómez-Tato A, Alvarez-Dios J, Calaza M, de Cal MC, Ballard D, Lareu MV, Carracedo A, The SNPforID Consortium. (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. Forensic Sci Int Genet 1: 273–80.

3. Lao O, Vallone PM, Coble MD, Diegoli TM, van Oven M, van der Gaag KJ, Pijpe J, de Knijff P, Kayser M. (2010) Hum Mutat 31: E1875–93.

4. Halder I, Shriver M, Thomas M, Fernandez JR, Frudakis T. (2008) A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. Hum Mutat 29: 648–58.

5. Fondevila M, Phillips C, Naveran N, Fernandez L, Cerezo M, Salas A, Carracedo A, Lareu MV. (2008) Case report: identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. Forensic Sci Int Genet 2: 212–8.

6. Phillips C, Prieto L, Fondevila M, Salas A, Gómez-Tato A, Alvarez-Dios J, Alonso A, Blanco-Verea A, Brión M, Montesino M, Carracedo A, Lareu MV. (2009) Ancestry analysis in the 11-M Madrid bomb attack investigation. PLoS One 4: e6583.

7. Pritchard JK, Stephens M, Donnelly P. (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–59.

8. Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N.

(2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27: 1713–24.

9. Phillips C. (2009) SNP databases. Methods Mol Biol 578: 43–71.

10. Amigo J, Phillips C, Lareu M, Carracedo A. (2008) The SNPforID browser: an online tool for query and display of frequency data from the SNPforID project. Int J Legal Med 122: 435–40.

11. Amigo J, Salas A, Phillips C, Carracedo A. (2008) SPSmart: adapting population based SNP genotype databases for fast and comprehensive web access. BMC Bioinformatics 9: 428.

12. Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, et al. (2002) A human genome diversity cell line panel. Science 296: 261–2.

13. Rosenberg NA, Li LM, Ward R, Pritchard JK. (2003) Informativeness of genetic markers for inference of ancestry. Am J Hum Genet 73: 1402–22.

14. Liu N, Chen L, Wang S, Oh C, Zhao H. (2005) Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. BMC Genet 30 Suppl 1: S26.

15. Haasl RJ, Payseur BA. (2011) Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. Heredity 106: 158–71.

16. Phillips C, Rodriguez A, Mosquera-Miguel A, Fondevila M, Porras-Hurtado L, Rondon F, Salas A, Carracedo A, Lareu MV. (2008) D9S1120, a simple STR with a common Native American-specific allele: forensic optimization, locus characterization and allele frequency studies. Forensic Sci Int Genet 3: 7–13.

17. Phillips C, Fernandez-Formoso L, Garcia-Magariños M, Porras L, Tvedebrink T, Amigo J, Fondevila M, Gomez-Tato A, Alvarez-Dios

J, Freire-Aradas A, Gomez-Carballa A, Mosquera-Miguel A, Carracedo A, Lareu MV. (2010) 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 (Epub ahead of print) doi:10.1016/j.fsigen.2010.02.003.

18. Falush D, Stephens M, Pritchard JK. (2003) Inference of population structure using mul-

tilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7: 574–8.

19. Falush D, Stephens M, Pritchard JK. (2007) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 4746–51.

20. Hubisz MJ, Falush D, Stephens M, Pritchard JK. (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resources 9: 1322–32.

Chapter 9

SNaPshot® Minisequencing Analysis of Multiple Ancestry-Informative Y-SNPs Using Capillary Electrophoresis

Maria Geppert and Lutz Roewer

Abstract

This protocol describes a strategy for analyzing phylogenetic Y-SNPs in a hierarchical multiplex assay by utilizing the SNaPshot® Multiplex System. Step by step, the protocol assists in the appropriate selection of SNPs, the primer design, the set up of PCR/SBE reactions as well as in the analysis of the results. Furthermore, a forensic approach is highlighted, in which the most probable ancestry of an unknown male DNA is inferred by the geographical distribution of the assigned Y-SNP haplogroup.

Key words: Multiplex PCR, Capillary electrophoresis, SNaPshot®, Minisequencing, Y-SNP, AIM

1. Introduction

For the analysis of SNPs, various technologies have been established over the last years (for a review see ref. (1)). SNaPshot® minisequencing became one of the most accepted techniques to type several SNPs of interest in a multiplex assay, because the method is uncomplicated, universally applicable (2–4), and can be accomplished with the standard devices that exist in a routine forensic lab (thermocycler, sequencer). The principle of the SNaPshot® minisequencing approach is the allele-specific extension of a primer that anneals adjacent to the position of the relevant SNP. Fluorescently labeled didesoxynucleotides (ddNTPs), which are known from the Sanger sequencing method, inhibit a further elongation, because of lacking the required hydroxyl group. Subsequently, the extended single base extension (SBE) primers which are different in length and color are analyzed by capillary electrophoresis (CE).

SNPs are frequent throughout the genome. For the Y chromosome alone, the dbSNP-Database comprises at the moment about

80,000 SNPs, although the number of validated markers is considerably smaller (approximately 600) (5). The mutation rate of SNPs is low in comparison to the STRs (about 10^{-8} vs. 10^{-3}). Combinations of SNP alleles located on the nonrecombinant portion of the Y chromosome usually pass unchanged from generation to generation and can easily be used to construct a unique phylogenetic tree (6). The succession of accumulating binary genetic markers as SNPs reveals a cascade of differentiation that randomly coincides with various population origin episodes, each with a specific temporal and geographical context (7). Thus, Y-SNPs are ancestry-informative markers (AIM) bearing information on the geographic ancestry of a DNA. The Y Chromosome Consortium (8) has published updated versions of the maximum parsimonious phylogenetic tree of human Y chromosomes and has proposed a universal nomenclature (http://ycc.biosci.arizona.edu). The major clades (haplogroups) are labeled with a capital letter (e.g., R) and subhaplogroups are designated alternately with numbers and small letters (e.g., R1b). Usually the terminal SNP is included to determine the branch unequivocally (R-M343 alias R1b). The most recent phylogenetic tree from 2008 consists of 311 haplogroups, which are defined by 600 SNPs (9). Nevertheless, the 1,000 Genomes Project Consortium has recently detected not less than 2,870 variable sites at the Y chromosome (74% novel), which support the already existing clades, but identified novel differentiation within established branches, e.g., in O2b (frequent in China, Japan) or R1b (Europe) (5).

The unmatched phylogenetic, geographical, and chronological resolution of the Y chromosome (10) provides insights into questions that are of interest to paleontologists, anthropologists, linguists, and historians and increasingly to forensic analysts. Each of the numerous population studies published weekly make our knowledge on geographical distributions of Y chromosomes more precise. Thus, the stage is set for an application of Y-SNPs in the field of forensics. The combination of a sensitive minisequencing method to analyze SNPs even from strongly degraded DNA (11) with population databases mapping the genetic differentiation among world populations (e.g., the Y Chromosome Haplotype Reference Database (YHRD); http://www.yhrd.org) helps to obtain information from biological material on the most probable geographical ancestry of missing persons or unknown suspects.

In this article, we will present a protocol for the analysis of Y-chromosomal SNPs with a hierarchical approach, in which the markers are typed according to their position in the phylogenetic tree of the YCC 2008, as it has been already proven for several population-specific Y-SNPs (12–14). Starting with a generic typing to assign the test sample to one of the major haplogroups, only the specific—and more geographically restricted—clades of the already detected major haplogroup have to be tested afterwards.
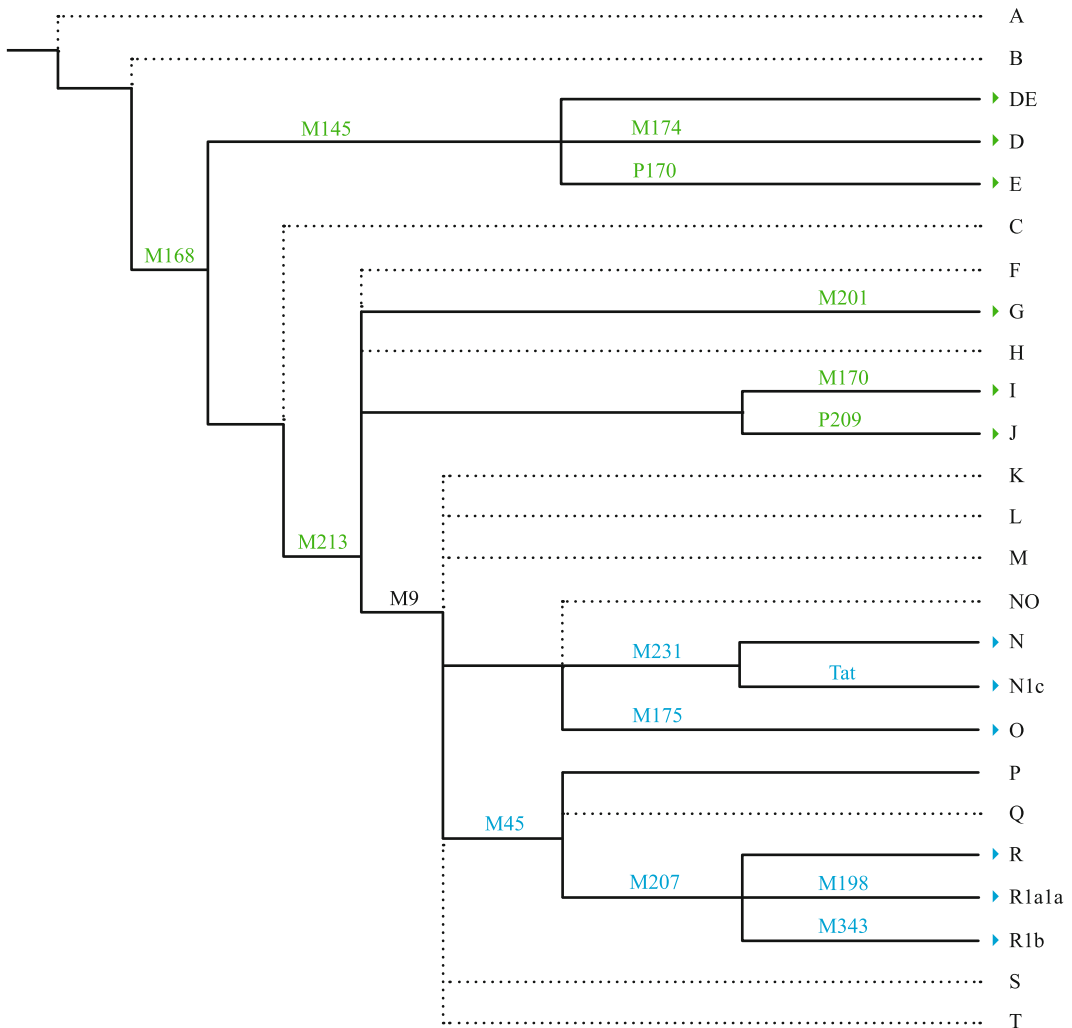
Fig. 1. Phylogenetic tree according to (9). *Solid lines* with marker indicate haplogroups, which can be typed by the two presented multiplexes. Multiplex I is marked in *green* and Multiplex II in *blue*. SNP M9 is marked *black*, because it is included in both assays. *Dotted lines* indicate haplogroups, which are not included in the multiplexes. The length of each branch has no significance.

In this way, the number of markers that have to be analyzed is limited to the minimum and the consumption of DNA material can be reduced. Two different multiplexes comprising 17 Y-SNPs, which were designed to detect the most common major haplogroups in samples from Germany, will serve as examples. Multiplex I covers the markers CT-M168, DE-M145, D-M174, E-P170, FT-M213, G-M201, I-M170, J-P209, and KT-M9 (see Fig. 1). Multiplex II includes the SNPs N-M231, N-Tat, O-M175, P-M45, R-M207, R-M198, and R-M343 (see Fig. 1) and for confirmation also SNP M9 (15). The result of the Multiplex I places the Y chromosome either in the part of the tree topology comprising the haplogroups

A-J or (if M9 is found derived) in the branches K-T. If the latter is the case, Multiplex II can be applied to define the clades downstream to M9 (see Fig. 1).

2. Materials

When working with chemicals always wear suitable protective cloth and gloves. Use ultrapure water to prepare all solutions and mastermixes. Avoid multiple thawing and freezing for all reagents which are stored at –20°C.

2.1. Selection of Y-SNPs, PCR-, and SBE Primer Design

1. dbSNP-Database (http://www.ncbi.nlm.nih.gov/projects/SNP/).
2. NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi).
3. Family Tree DNA Y Chromosome Browser (http://ymap.ftdna.com/).
4. Primer Quest (http://eu.idtdna.com/Scitools/Applications/Primerquest/).
5. Primer3Plus (http://www.primer3plus.com).
6. Oligocalculator (http://www.basic.northwestern.edu/biotools/oligocalc.html).
7. UCSC BLAT Search Genome (http://genome.ucsc.edu/).
8. UCSC In-Silico PCR (http://genome.ucsc.edu/).

2.2. PCR Reactions

1. AmpliTaq Gold® DNA Polymerase (Applied Biosystems, Inc., Foster City, CA, USA).
2. GeneAmp® 10× PCR Buffer (Applied Biosystems).
3. dNTPs.
4. $MgCl_2$.
5. Multiplex PCR Kit (Qiagen®, Hilden, Germany).

2.3. Agarose Gel Components

1. Agarose.
2. Tris-Borate-EDTA (TBE) buffer.
3. Ethidium bromide.
4. 3× Concentrate loading buffer: 10% glycerol, 2.5% Ficoll 400, 0.025% bromophenol blue, 0.025% xylene cyanol. For a 3× concentrated 10 mL solution loading buffer weight 25 mg bromophenol blue, 25 mg xylene cyanol, 1 mL glycerol, 0.25 g Ficoll 400, 8.8 mL ultrapure water.
5. Hyperladder™ V (Bioline, Luckenwalde, Germany).

2.4. PCR Cleanup, SNaPshot® Single Base Extension Reaction, and SBE Reaction Cleanup	1. It is ExoSAP-IT® (USB®—Affymetrix, Cleveland, OH, USA). 2. ABI Prism® SNaPshot® Multiplex Kit (Applied Biosystems). 3. Shrimp Alkaline Phosphatase (SAP, USB®).
2.5. Capillary Electrophoresis	1. ABI Prism® Genetic Analyzer 3130 or 3130xl with 36-cm capillaries (Applied Biosystems). 2. SNaPshot® Matrix Standard Set DS-02 (dR110, dRGG, dTAMRA™, dROX™, LIZ®). 3. Performance Optimized Polymer (POP-4 or POP-7; Applied Biosystems). 4. GeneScan™ LIZ™ 120 size standard (Applied Biosystems). 5. Hi-Di™ Formamide (Applied Biosystems).
2.6. Data Analysis	1. Genemapper® ID 3.2.1. 2. Y Chromosome STR Haplotype Reference Database (YHRD) http://www.yhrd.org.

3. Methods

3.1. Selection of Y-SNPs, PCR-, and SBE Primer Design	1. The selection of Y-SNPs has to be made with respect to the population(s) under study, e.g., branch R-specific SNPs for the study of European or branch O-specific SNPs for East Asian lineages. Already published studies help to determine the most common population-specific haplogroups and the appropriate markers. For samples from Germany, the haplogroups DE, E, G, I, J, N, O, R1a, and R1b are most common (16–18) for which the relevant SNP markers should be arranged in multiplex PCRs (see Fig. 1). 2. Subsequently, the primers have to be designed to run in a multiplex assay (see Note 1). To find suitable primers, pick the sequences, which are 200 bp up- and downstream of the relevant SNPs. While having a SNP with Reference SNP ID (rs number), the search using the dbSNP-Database is quick and uncomplicated. Obtaining the results for SNPs that are not registered in the NCBI is more difficult. An alternative approach for those SNPs is to BLAST primer sequences of the SNPs of interest, which have already been published, or to interrogate the Family Tree DNA Y Chromosome Browser. Only those markers should be selected which are well studied and known to have no paralogous sequence variants (16, 17). In our opinion, it gives more reliable results, if the relevant SNPs are distributed to more than one but hierarchically

applicable multiplex instead of designing a very large multiplex covering all SNPs.

3. Amplicons of the SNPs should be as short as possible to utilize the full potential of the SNPs especially to amplify degraded DNA. The size of the primers should be between 20 and 25 nucleotides depending on the sequence, because the melting temperature should be theoretical between 58 and 62°C at a salt concentration of 180 mM by considering that the GC/AT content is more or less equal. The variance of melting temperatures within a multiplex assay should be kept small. Amplicons of the different SNPs should differ in size at least 3–5 nucleotides to allow for a clear electrophoretic separation.

4. Primers should be checked for possible extendable hairpin structures (loops) within each primer and for primer dimers. In a multiplex assay with about ten markers, it is not possible to avoid all theoretical primer inter- and intracontacts, but the most precarious ones. Particularly those should be excluded which lead to free extendable 3′ ends. Primer Quest, Primer3Plus, and OligoCalulator were used to design and check the primers for secondary structures. NCBI BLAST, UCSC BLAT Search Genome, and UCSC In-Silico PCR were used to check binding sites of the primers in the genome to assure that all PCR primers are specific for the Y chromosome.

5. The design of SBE primers can be conducted with the same software as it has been used for the PCR primers. Depending on the flanking sequence of the SNP, primers can be designed to anneal to the forward or reverse strand (see Note 2). In the capillary electrophoresis, the single SNPs are differentiated by the color of the added nucleotide and by the length of the primer. To separate several SNPs of a multiplex reaction, both variables have to be used. SBE primers are extended to a certain length with polynucleotide tails or an unspecific nonhuman sequence (18). The latter is more stable and less affected by degradation (see Note 3). The melting temperature of the SBE primers should be at least 50°C, which corresponds with a length of the specific part of 17–18 nucleotides. Aiming to separate 8–12 SNPs in a range of ~60 nucleotides, the distance between SBE primers and their order is crucial. While short primers (17–30 nucleotides) should differ between 5 and 7 nucleotides in length, longer primers can have only 4 nucleotides distance to each other, because the resolution of the CE increases with the length of the fragments. Primers can be set to the same length, if the corresponding SNP mutations do not share the same nucleotides (e.g., G/C and A/T), because these can be easily distinguished by the specific colors. SBE primers should be HPLC purified (19).

3.2. PCR Reactions

1. Testing primers by singleplex PCR
   For testing the functionality of the designed primers, each pair should be amplified in a singleplex reaction by preparing a positive control of a well-known DNA sample, a female control, and a negative control. In a final volume of 12.5 µL, mix 1 U Taq Gold, 0.25 mM dNTPs, 1× PCR buffer, 1 mM MgCl, 0.5 µM of each primer, and 1 ng DNA and fill up with ultra-pure water.
   Cycle the PCR with the following protocol: 95°C for 10 min followed by 8 cycles of 95°C for 30 s, 60°C for 30 s, 72°C for 30 s, 26 cycles of 95°C for 30 s, 58°C for 30 s, 72°C for 30 s, and a final extension of 72°C for 5 min.

2. SNaPshot® multiplex PCR
   The multiplex PCR reaction is performed in a total volume of 6 µL. Prepare for one reaction a mix of primers (see Table 1) with 1× concentrated Multiplex PCR Mastermix. Add 1 µL genomic DNA of a single individual and ultrapure water to obtain the final volume. The amount of 1 ng/reaction was tested to work well. Vortex briefly and centrifuge the content to the bottom before thermocycling. Always include a positive and a negative control into your calculation and take pipetting mistakes into account (see Note 4).
   Cycle the PCR with the following protocol: 95°C for 15 min followed by 32 cycles of 94°C for 30 s, 60°C for 90 s, 72°C for 60 s, and a final extension of 72°C for 10 min.

3.3. Gel Electrophoresis for Singleplex and Multiplex PCR Control

1. For the electrophoretic control of the singleplex amplification, a minigel of 1.4% agarose and 0.035% Ethidium bromide can be used. Run 2 µL of the sample with 4 µL of the 1× concentrated loading buffer for 20 min at 100 V. Visualize the bands under UV transillumination.

2. For the electrophoretic separation of multiplex PCR products, prepare a gel (20 cm length) containing 2.5–3% agarose and 0.035% Ethidium bromide. Run 2 µL of the sample with 4 µL of the 1× concentrated loading buffer for approximately 2 h at 100 V or until a good resolution is reached. Visualize the bands under UV transillumination (see Note 5; Fig. 2).

3.4. PCR Cleanup

Before SBE, clean up the PCR product with ExoSAP-IT® (USB) to remove PCR primers and not incorporated dNTPs, which would participate in subsequent reactions. Because of the high glycerol concentration, vortex briefly and incubate the reaction 60 min at 37°C followed by 15 min at 75°C to inactivate the enzymes (19).

Table 1
PCR primer information grouped by multiplexes and sorted by the according branch

| SNP	Branch	Forward primer (5′–3′)	bp	Reverse primer (5′–3′)	bp	Amplicon (bp)	SNP	Concentration (µM)	
M168	CT	TGTTTTGCAGAGAGCTTGGA	20	TGACTGTTCAGTTTATTCCACAAA	25	150	C/T	0.22	Multiplex I
M145	DE	GCATACTTGCCTCCAGACT	20	CCAGGAGCTCACAGTCACAA	20	173	G/A	0.22	
M174	D	TTCTCCGTCACAGCAAAATG	21	ATGCAAAAGGAGAAGGACAAGA	22	186	T/C	0.22	
P170	E	CCTCCTGTGCCTCTTTCAGA	20	ACAGCAGCAAGCAGGTCTTT	20	243	G/A	0.5	
M213	FT	GGCCATATAAAAGGCAGCA	20	TGAATGGCAAATTGATTCCA	20	208	T/C	0.34	
M201	G	TATGCAITTGTTGAGTATATGTC	23	GTTCTGAATGAAAGTTCAAACG	22	326	G/T	0.22	
M170	I	TTGCAGCTCTTATTAAGTTATG	21	TAAACTTGCCAATTACTTTCAAC	23	167	A/C	0.5	
P209	J	TAGTCAAGAAGGCAGAATAGC	21	GTAAGGTATATTACTGAATCCAC	23	299	T/C	0.5	
M9	K-T	GCAGCATATAAAACTTTCAGG	21	AAAACCTAACTTTGCTCAAGC	21	340	C/G	0.5	
M9	KT	GCAGCATATAAAACTTTCAGG	21	AAAACCTAACTTTGCTCAAGC	21	340	C/G	0.33	Multiplex II
M231	N	CCTATTATCCTGGAAAATGTGG	22	GAATGGTGGCCAGAGTCTTT	20	219	G/A	0.33	
Tat	N1c	GACTCTGAGTGTAGACTTGTGA	22	GAAGGTGCCGTAAAAGTGTGAA	22	112	T/C	0.33	
M175	O	TAAACTCTCTGAATCAGGCACATG	24	CTACTGATACCTTTGTTTCTG	21	71	–5 bp (T/A)	0.33	
M45	P	GAGAGAGGATATCAAAAATTGG	22	TAGCTTACAACACAAGGATTC	21	229	G/A	0.33	
M207	R	CTATGGGGCAAATGTAAGTC	20	TGAAGGAAAAGTGGAGTCTG	20	129	A/G	0.33	
M198	R1a1a	ATTCCAGTCATGATGAGGTGG	21	ATGCCGTTTGCCTAGGTTAG	21	138	C/T	0.33	
M343	R1b	TCTGATTCGCACAAGGCTC	19	CACCTTTGTCCTCTGCTC	19	194	C/A	0.33	
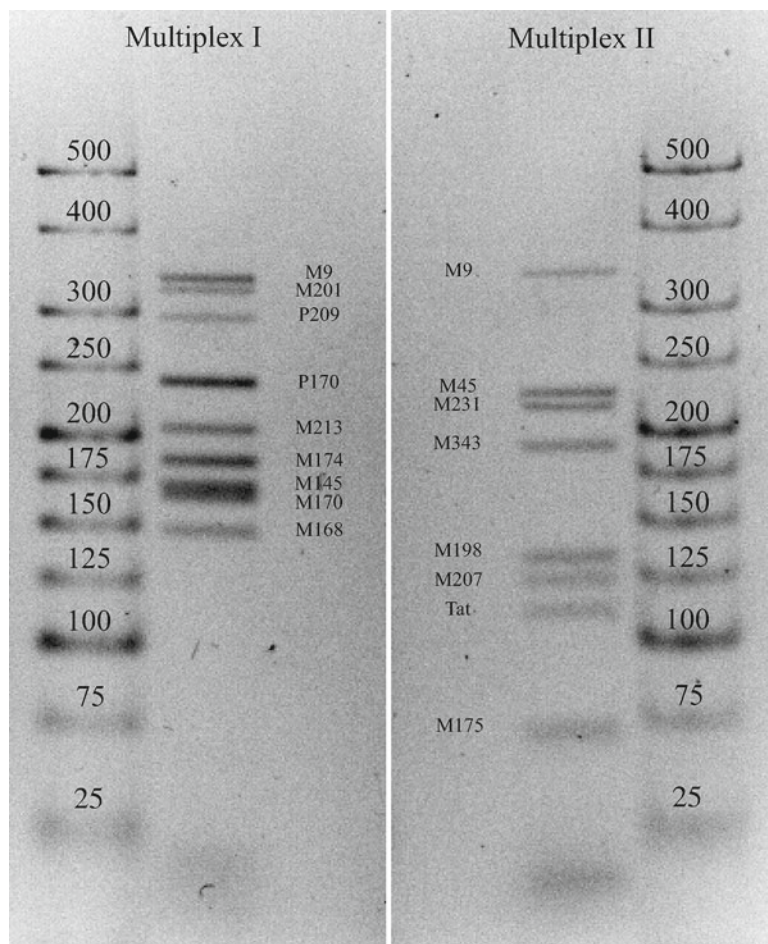
Fig. 2. Control of multiplex PCRs by gel electrophoresis (inverted image). Hyperladder® V was used as a size marker.

3.5. SBE Reaction

1. Singleplex SBE reaction for control
   Before running all SBE primers in one analysis, we recommend to run each SBE primer in a single extension reaction. Therefore, mix 1.5 µL SNaPshot® Multiplex Ready Reaction Mix (Applied Biosystems), 1.5 µL of the cleaned up singleplex or multiplex PCR product, and SBE primer (see Table 2). Add ultrapure water to a total volume of 5 µL (see Note 6). Thermocycling conditions are described in the subsequent section of the multiplex SBE reaction.

2. Multiplex SBE reaction
   The multiplex SBE reaction is performed in a final volume of 5 µL, consisting of 1.5 µL SNaPshot® Multiplex Ready Reaction Mix (Applied Biosystems), 1.5 µL cleaned up PCR product, SBE primers (see Table 2), and ultrapure water (see Note 7). Setting up the SBE reaction should be performed on ice or in a frozen rack and it should be finished within 20 min to avoid a higher background in the capillary electrophoresis (19).

Table 2

Sequencing primer information grouped by multiplexes and sorted by primer size

SNP	Primer sequence (5′–3′) target-specific sequence *black*, tail sequence *blue*	Size (bp)	Orientation typing	Concentration (µM)	
M201	AATAATCCAGTATCAACTGAGG	22	Forward	0.09	Multiplex I
M174	CTGACAAGCACCCCTCACTTCTGCACT	27	Reverse	0.09	
P209	AAGTCTGACAAACTTCATCTCTGAGTCATT	30	Forward	0.23	
M213	AAGTCTGACAAGAACTTAAAACATCTCGTTAC	32	Reverse	0.09	
P170	CGTCGTGAAAGTCTGACAATTTCCTTTGGCAAACTGA	37	Forward	0.14	
M9	GGTGCCACGTCGTGAAAGTCTGACAAGCCTAAGATGGTTGAAT	43	Forward	0.27	
M145	CTAGGTGCCACGTCGTGAAAGTCTGACAAGACACCAGAAAGAAAGGC	47	Forward	0.09	
M168	AACTAGGTGCCACGTCGTGAAAGTCTGACAAGTATGTGTTGGAGGTGAGT	50	Forward	0.09	
M170	AACTAGGTGCCACGTCGTGAAAGTCTGA CAAATTTACTTAAAAATCATTGTTC	53	Forward	0.27	
M207	ATGTAAGTCAAGCAAGAAATTTA	23	Forward	0.18	Multiplex II
M45	TCTGACAACAGAAGGAGCTTTTGC	25	Reverse	0.18	
M198	GACAATGTACTTAAATTAACCTAAAAGA	28	Reverse	0.18	
M175	AAAGTCTGACAACACATGCCTTCTCACTTCTC	32	Forward	0.18	
M231	TCGTGAAAGTCTGACAATTACTGTTTCTACTGCTTTC	37	Forward	0.18	
M9	GGTGCCACGTCGTGAAAGTCTGACAAGCCTAAGATGGTTGAAT	43	Forward	0.18	
Tat	GCCACGTCGTGAAAGTCTGACAACTCTGAAATATTAAATTAAAACAAC	48	Reverse	0.18	
M343	ACTAAACTAGGTGCCACGTCGTGAAAGTC TGACAAGAGTGCCCTCGTGTTCCA	53	Forward	0.18	

Cycle the SBE reaction with following conditions: 96°C for 10 s, 50°C for 5 s and 60°C for 30 s, for 25 cycles (19).

3.6. SBE Reaction Cleanup

Not incorporated ddNTPs have to be removed to avoid unspecific dye blobs during capillary electrophoresis. Therefore, the SBE reaction is cleaned up with 1 U of shrimp alkaline phosphatase (USB) by incubating the thoroughly mixed reaction at 37°C for 60 min and inactivating the enzyme at 85°C for 15 min (19).

3.7. Capillary Electrophoresis (Avant 3130/3130xl)

To run the SNaPshot® reaction, mix 15 µL Hi-Di formamide and 0.1 µL Gene Scan 120 Liz/sample and apply 15 µL of the mix to each well. Add 1 µL of the purified SBE sample and mix thoroughly (see Note 8). Denaturation prior to running the analyzer is not necessary. Default run module for SNPs was used as supplied for 36 cm capillaries either with POP-4 or POP-7.

3.8. Data Analysis

1. Genemapper® analysis
   For analyzing the SNaPshot® results, it is useful to create a panel with the appropriate bin set of a multiplex to allow for an automated allele calling (20). The singleplex SBE runs should be used for setting up the bins.

2. Data interpretation
   After evaluating the SNP alleles, the haplogroup of a test sample can be determined. An example is shown in Fig. 3. In the first multiplex (see Fig. 3a), only the generic markers CT-M168, FT-M213, and KT-M9 are mutated (derived state), thus a further testing was required. In Multiplex II (see Fig. 3b), the SNPs P-M45, R-M207, and R-M343 show the derived state, which leads to the assignment of the sample to haplogroup R1b with the final marker M343. Searching the actual YHRD release 34, 707 R1b chromosomes, out of 6,246 reference samples with haplogroup information, can be found in 31 of total 88 worldwide populations. The geographical distribution indicates the highest frequency of R-M343 in the Western and Central part of Europe (compare Fig. 4), which is in agreement with the published literature (10, 21–23). The frequent occurrence in South America can lead back to European emigration. The most probable geographic origin of the ancestors of the tested man is therefore Europe.

4. Notes

1. Be careful with using primers or complete assays, which have already been published. Always check for correctness of the primer sequences by using BLAST and if the primers are consistent with your requirements. Check the primer sequences, amplicon length, annealing temperature, and characteristics of the SNP.
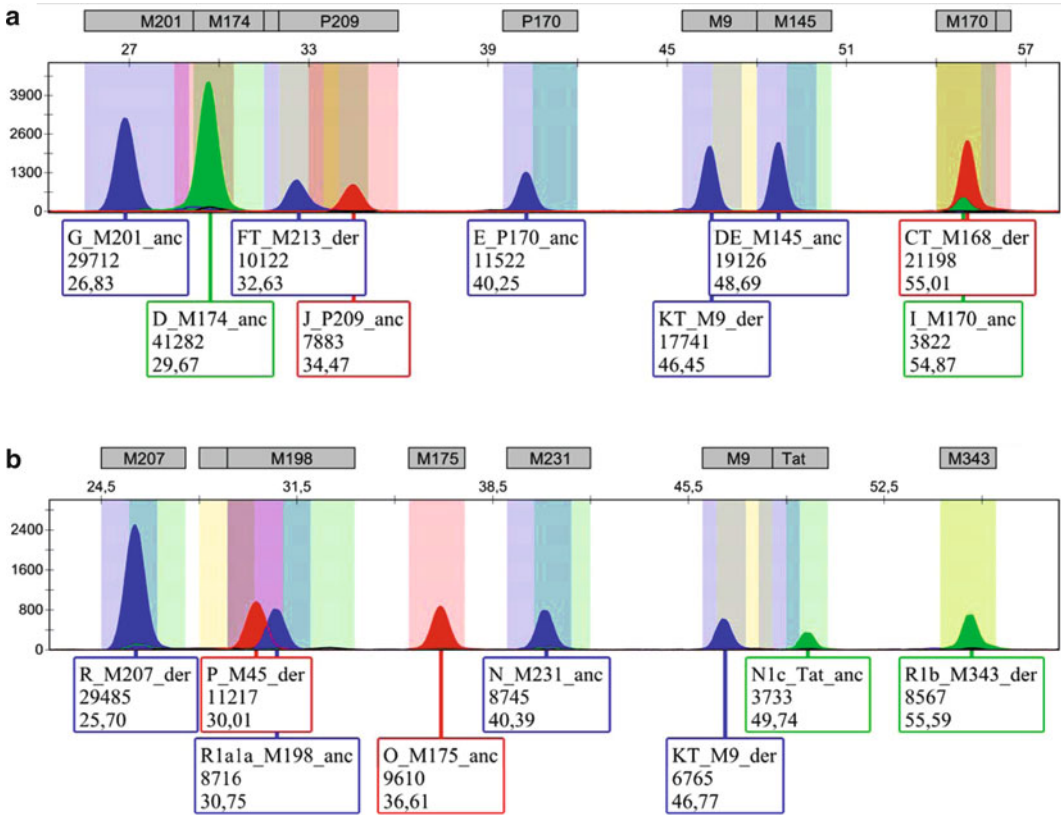
Fig. 3. Electropherograms of the test sample assigned to haplogroup R-M343. (**a**) Multiplex I shows three SNPs with derived state—M168, M213, M9. (**b**) Multiplex II shows four markers with derived state—M9 (like in Multiplex I), M45, M207, and M343. The *gray boxes* above the peaks indicate the size range of both allele-specific primers of a SNP according to the bins. *Peak colors* indicate the didesoxynucleotide (ddNTP) which annealed to the according SBE primer. *Colors* of the bins match with the allele-specific nucleotide they are calling. Labels give information about the branch that the SNP is defining, the name of the SNP, the allele status of the SNP, the RFU signal, and the length of the primer.

2. Consider that with changing the direction of the sequencing primer, the nucleotides of the ancestral and derived status of the SNP are changing as well.

3. In our experiments, poly (dC) tails showed a tendency to degrade rapidly and generate incorrect elongations.

4. For setting up a new assay, primers should be tested with equal concentrations (e.g., 0.3 µM/primer). Afterwards, the primer concentrations should be adjusted to their performance, aiming that all amplicons show similar intensity in the gel electrophoresis.

5. If there are samples without visible amplification products, nevertheless it is worthwhile to go on in the protocol. The electrophoresis is not sensitive enough to image weak products, that is why in some of the cases you will gain a SNP profile.
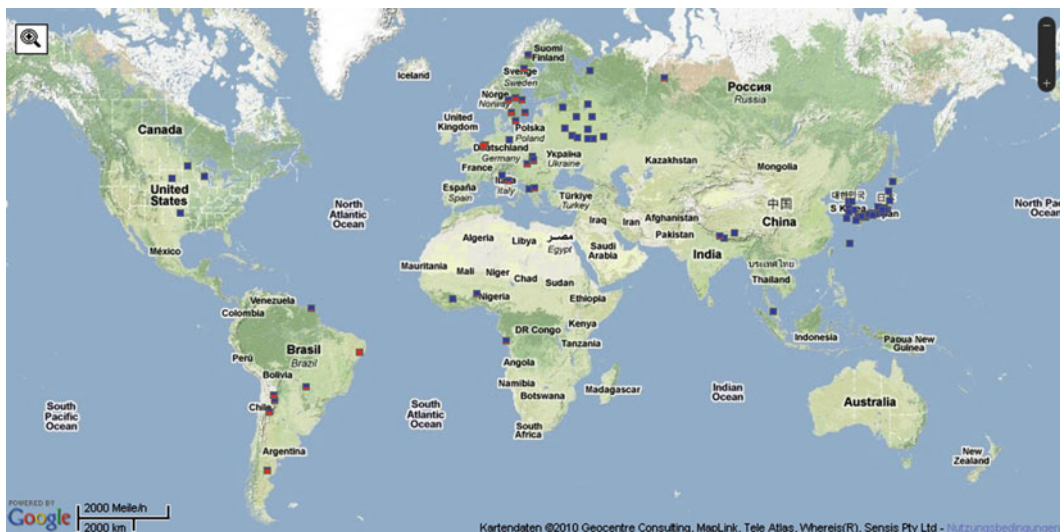
Fig. 4. Search result and map of the geographical distribution of haplogroup R-M343 (R1b) in the Y Chromosome Haplotype Reference Database (YHRD) (http://www.yhrd.org). *Red boxes* depict geopositions where the questioned haplogroup matched. The proportion of *red/blue* in a single box correlates with the relative frequency. *Blue boxes* indicate geopositions, where haplotypes with SNP information are available, but the questioned haplogroup does not occur.

6. Using the SBE singleplex analyses, the factual length of the primers can be checked in CE and compared to the theoretical one. Especially short SBE primers between 17 and 30 nucleotides can show enormous differences.

7. The first multiplex SBE attempt with all primers should be done with equal primer concentrations (e.g., 0.2 μM). Subsequently, the concentrations can be finally adapted. A well-balanced multiplex increases the chance to gain results even from compromised samples.

8. Depending on the sensitivity of the Genetic Analyzer, the required quantity of the size standard and the sample have to be adjusted. If peaks in CE are higher than 5,000 RFU, cleaned up SBE products can be diluted to allow for a clear determination of alleles.

References

1. Sobrino, B., and Carracedo, A. (2005) SNP typing in forensic genetics: a review, *Methods Mol Biol 297*, 107–126.

2. Sanchez, J. J., Phillips, C., Borsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, D., Schneider, P. M., Carracedo, A., and Morling, N. (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification, *Electrophoresis 27*, 1713–1724.

3. Brandstatter, A., Parsons, T. J., and Parson, W. (2003) Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups, *Int J Legal Med 117*, 291–298.

4. Walsh, S., Liu, F., Ballantyne, K. N., van Oven, M., Lao, O., and 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*(3), 170–80.

5. A map of human genome variation from population-scale sequencing, *Nature 467*, 1061–1073.

6. Jobling, M. A. (2001) Y-chromosomal SNP haplotype diversity in forensic analysis, *Forensic Sci Int 118*, 158–162.

7. Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Mirazon Lahr, M., Foley, R. A., Oefner, P. J., and Cavalli-Sforza, L. L. (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations, *Ann Hum Genet 65*, 43–62.

8. (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups, *Genome Res 12*, 339–348.

9. Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L., and Hammer, M. F. (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, *Genome Res 18*, 830–838.

10. Jobling, M. A., and Tyler-Smith, C. (2003) The human Y chromosome: an evolutionary marker comes of age, *Nat Rev Genet 4*, 598–612.

11. Bouakaze, C., Keyser, C., Amory, S., Crubezy, E., and Ludes, B. (2007) First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA, *Int J Legal Med 121*, 493–499.

12. Brion, M., Sobrino, B., Blanco-Verea, A., Lareu, M. V., and Carracedo, A. (2005) Hierarchical analysis of 30 Y-chromosome SNPs in European populations, *Int J Legal Med 119*, 10–15.

13. Sanchez, J. J., Borsting, C., Hallenberg, C., Buchard, A., Hernandez, A., and Morling, N. (2003) Multiplex PCR and minisequencing of SNPs--a model with 35 Y chromosome SNPs, *Forensic Sci Int 137*, 74–84.

14. Onofri, V., Alessandrini, F., Turchi, C., Pesaresi, M., Buscemi, L., and Tagliabracci, A. (2006) Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs, *Forensic Sci Int 157*, 23–35.

15. King, T. E., and Jobling, M. A. (2009) Founders, drift, and infidelity: the relationship between Y chromosome diversity and patrilineal surnames, *Mol Biol Evol 26*, 1093–1102.

16. Sanchez, J. J., Brion, M., Parson, W., Blanco-Verea, A. J., Borsting, C., Lareu, M., Niederstatter, H., Oberacher, H., Morling, N., and Carracedo, A. (2004) Duplications of the Y-chromosome specific loci P25 and 92R7 and forensic implications, *Forensic Sci Int 140*, 241–250.

17. Adams, S. M., King, T. E., Bosch, E., and Jobling, M. A. (2006) The case of the unreliable SNP: recurrent back-mutation of Y-chromosomal marker P25 through gene conversion, *Forensic Sci Int 159*, 14–20.

18. Lindblad-Toh, K., Winchester, E., Daly, M. J., Wang, D. G., Hirschhorn, J. N., Laviolette, J. P., Ardlie, K., Reich, D. E., Robinson, E., Sklar, P., Shah, N., Thomas, D., Fan, J. B., Gingeras, T., Warrington, J., Patil, N., Hudson, T. J., and Lander, E. S. (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse, *Nat Genet 24*, 381–386.

19. Protocol: ABI PRISM® SNaPshot® Multiplex System http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_041203.pdf.

20. SNaPshot® Kit Analysis - Getting Started Guide, http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/general-documents/cms_042037.pdf.

21. Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A. S., Cavalli-Sforza, L. L., and Underhill, P. A. (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective, *Science 290*, 1155–1159.

22. Rosser, Z. H., Zerjal, T., Hurles, M. E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D. G., Brede, G., Cooper, G., Corte-Real, H. B., de Knijff, P., Decorte, R., Dubrova, Y. E., Evgrafov, O., Gilissen, A., Glisic, S., Golge, M., Hill, E. W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S. A., Krumina, A., Kucinskas, V., Lavinha, J., Livshits, L. A., Malaspina, P., Maria, S., McElreavey, K., Meitinger, T. A., Mikelsaar, A. V., Mitchell, R. J., Nafa, K., Nicholson, J., Norby, S., Pandya, A., Parik, J., Patsalis, P. C., Pereira, L., Peterlin, B., Pielberg, G., Prata, M. J., Previdere, C., Roewer, L., Rootsi, S., Rubinsztein, D. C., Saillard, J., Santos, F. R., Stefanescu, G., Sykes, B. C., Tolun, A., Villems, R., Tyler-Smith, C., and Jobling, M. A. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language, *Am J Hum Genet 67*, 1526–1543.

23. Myres, N. M., Rootsi, S., Lin, A. A., Jarve, M., King, R. J., Kutuev, I., Cabrera, V. M., Khusnutdinova, E. K., Pshenichnov, A., Yunusbayev, B., Balanovsky, O., Balanovska, E., Rudan, P., Baldovic, M., Herrera, R. J., Chiaroni, J., Di Cristofaro, J., Villems, R., Kivisild, T., and Underhill, P. A. (2011) A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe, *Eur J Hum Genet. 19*(1), 95–101.

Chapter 10

Capillary Electrophoresis of 38 Noncoding Biallelic Mini-Indels for Degraded Samples and as Complementary Tool in Paternity Testing

Rui Pereira and Leonor Gusmão

Abstract

This work describes the main advantages and the steps involved in the optimization of a multiplex system able to characterize 38 noncoding biallelic *Insertion Deletion Polymorphisms* (Indels). With this methodology, all markers are amplified in a single PCR, using short amplicons (up to 160 bp) in order to improve its performance in degraded samples. Alleles are easily detected using capillary electrophoresis.

The Indel multiplex typing strategy here described has the same desirable characteristics of forensic SNP assays, including genetic markers (a) with low mutation rates, increasing their usefulness in some kinship cases where few or single incompatibilities can be explained by mutation, and (b) that can be typed using a short amplicon strategy, increasing their usefulness in cases where degraded samples are available. Moreover, this approach uses simple and well-established methodologies already applied in forensic STR assays.

Key words: Insertion deletion polymorphism, Indel, DIP, Multiplex PCR, Human identification, Paternity testing, Forensic genetics

1. Introduction

The human DNA sequence encompasses distinct types of variation, each presenting desirable characteristics to address different biological questions. For example, STRs present high degrees of length polymorphism in human populations and are predominantly multiallelic. Because they are highly diverse, this type of genetic markers is very informative and widely used in population and forensic genetics. Conversely, mutational events such as nucleotide substitutions, insertions, or deletions occur with a very low frequency, possibly only once in the recent evolutionary past of modern humans. By their nature, these mutations create polymorphisms

with two allelic states (ancestral and mutant) and are called binary or biallelic polymorphisms. These genetic markers present lower genetic diversity than STRs, which is a disadvantage regarding forensic applications, but they also have several desirable features as genetic markers in this field. The usefulness of SNPs in forensics is well patent in many publications (e.g., see ref. (1–6)) or in other chapters of this book, and here we dedicate special attention to insertion deletion (Indel) polymorphisms, particularly in the analysis of degraded samples and as complementary tool in paternity testing.

1.1. Insertion Deletion Polymorphisms

The studies by Weber et al. (7) and Mills et al. (8) contributed fundamentally to our knowledge on Indel polymorphisms and highlighted the utility of Indels for most genetic studies given their abundance in the genome and ease of analysis. In 2002, Weber et al. (7) identified and characterized ~2,000 human biallelic Indels with different length variation and measured the allele frequencies in Europeans, Africans, Japanese, and Native Americans. In association with this work, Marshfield Clinic created an Insertion/Deletion Polymorphism online database (http://www.marshfieldclinic.org/mgs) including information regarding the type of polymorphism (biallelic or multiallelic), allele length differences, amplification primers, position in the chromosome, as well as allele frequencies in different population groups. Later in 2006, an initial map of Indel variation in the human genome was reported (8), comprising more than 415,000 polymorphisms. According to Mills et al. (8), insertion/deletions represent ~16% of all human DNA polymorphisms and are widely spread throughout the genome, with an average density of one Indel per 7.2 kb. Moreover, a particular class of apparently random DNA sequences comprises about 41% of all Indels and shows a wide range of allele length variation, from 2 up to ~10 kb, with nearly 99% being under 100 bp (8). Of great interest as genetic markers, these small length Indels are abundant in the human genome and amenable to simple analysis through PCR amplification and electrophoresis, or using methodologies already developed for SNPs (7–9).

Another important aspect is that Indel data from these and other studies have been compiled in the free access online database dbSNP from NCBI (under the MARSHFIELD and DEVINE_LAB submitter handle, respectively, for refs. (7) and (8)), which enables a much easier access to the available information.

In the last few years, there was an increasing interest of the scientific community in the use of Indel polymorphisms in different research projects, as for example when studying the genetic structure of human populations (10–12), as ancestry informative markers (9, 13), in X-chromosome specific panels (14–16), and more recently in human identification studies (17, 18).

1.2. Insertion Deletion Polymorphisms as Genetic Markers in Forensics

Considering the above-mentioned aspects, it becomes clear that some types of Indels combine interesting characteristics of other genetic markers commonly used in forensics (STRs and SNPs):

– Indels are abundant and widely distributed along the genome.

– The genotyping of small Indels is relatively easy and inexpensive with a simple PCR and electrophoresis approach.

– Small Indels can be analyzed in short amplicons, improving successful amplification of degraded DNA and, in the other hand, opening perspectives for good multiplexing levels.

– Small Indels are also suitable for automation and analysis with high-throughput technologies.

– Indels show very low mutation rates (occurring at low frequency, the polymorphism most likely derives from a single nonrecurrent mutation event in the evolutionary time of modern humans).

– Many Indels have significant divergences in allele frequencies among geographically separated population groups (therefore, they can be used as Ancestry Informative Markers).

1.3. Mini-Indels in the Analysis of Degraded Samples

Some of these characteristics are the basis behind the recent interest in Indels for forensic applications. The most relevant is the possibility of analysis in much shorter amplicons when compared with standard STRs. This was definitely the main reason for introducing SNPs in forensic practice, aiming to overcome the major drawback of STR analysis: its weak performance with highly degraded samples due to the long template DNA needed for amplification. In this work, we focus on mini-Indels with small length variation between alleles (2–5 bp), which likewise SNPs can be analyzed in very short amplicons (under 160 bp) thus improving the amplification success with challenging DNA (17, 19). The advantage of this approach is that the detection of alleles is accomplished simply by fragment separation with capillary electrophoresis, conversely to laborious SNP typing protocols. In summary, Indels ideally allow combining a short amplicon strategy (see Fig. 1a) with a simple analytic workflow (see Fig. 1b).

1.4. Indels as Complementary Tool in Paternity Testing

In paternity or other biological relationship testing situations, it can happen that few mendelian transmission incompatibilities are observed for STRs; and it is necessary to compare the probability of the different hypothesis assuming exclusion vs. the occurrence of a mutation. Nevertheless, after that assessment, the conclusion can sometimes remain unclear or need stronger support. Due to the high mutation rates of most STRs, this is not an uncommon scenario in paternity testing. In such cases, the complementary study of a large set of Indels or SNPs showing low mutation rates and therefore being more stable than microsatellites can help circumvent difficulties (4, 5, 19). Characterizing a high number of markers can
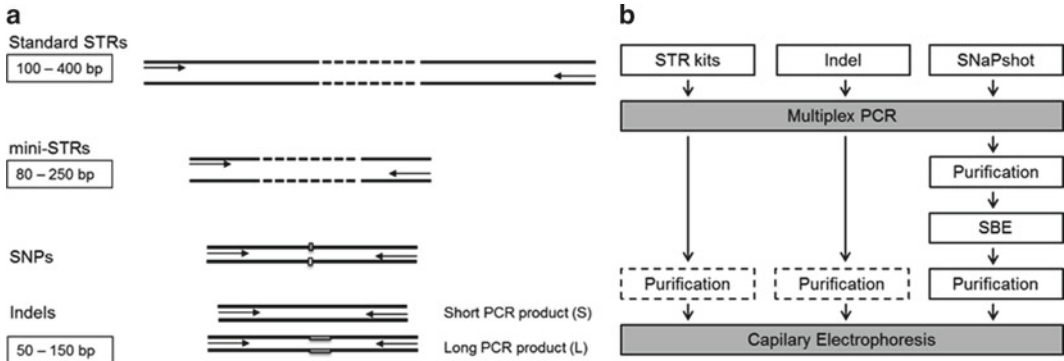
Fig. 1. Insertion deletion (Indel) polymorphism analysis in short amplicon and simple workflow: (**a**) Indels, like SNPs, are amenable to analysis in much shorter amplicons than standard STRs or even the re-engineered mini-STRs; (**b**) Indels, like STRs, can be easily typed with a PCR and capilary electrophoresis, conversely to the laborious protocol most widely used in forensic SNPs typing (SNaPshot; Applied Biosystems).

unveil further transmission incompatibilities, favoring the exclusion hypothesis or, in case of inclusion, increase the paternity index values to more confident levels supporting paternity (see Note 1).

1.5. General Considerations About the Indel Assay

The multiplex described in the present work includes 38 Indels that were previously found to be polymorphic and very informative in human population groups from Africa, Europe, and East Asia, with an average heterozygosity of 0.45 for the set. No statistically significant deviations from Hardy–Weinberg expectations were found in the studied populations and linkage disequilibrium tests also did not reveal statistically significant allelic association between the different loci pairs in the same populations (17).

Regarding forensic efficiency parameters, the 38 Indel set proved to be highly informative in African, European, and East Asian populations, with accumulated random match probabilities ranging in orders of magnitude from 10^{-14} to 10^{-15} and accumulated discrimination power reaching 99.999999999994% in the less diverse group. Despite being less powerful than standard STR kits (when full profiles are obtained), the Indel set appears to be enough informative in most forensic cases. Even more, it exceeds the efficiency parameters of the mini-STR kit Minifiler (Applied Biosystems), which likewise the mini-Indel multiplex is indicated to the analysis of highly degraded samples (17).

The analytical workflow of the Indel assay is very simple and, in comparison to the most widely available SNP typing methods, reduces considerably the steps needed to genotype a large set of informative biallelic markers for identification purposes. This is advantageous in terms of time and cost-effectiveness of the assay, and more importantly, it minimizes manipulation and risks of contamination or samples' mix up. Furthermore, the use of the same methodology as in standard STR genotyping makes this assay easy to implement in any forensic laboratory.

2. Materials

1. A number of free access online databases and computer programs were used in the selection of Indel markers and development of the multiplex here described (see Table 1 for details).

Table 1
Online databases and computer programs used in developing the Indel multiplex

Diallelic insertion/deletion polymorphisms database from Marshfield Clinic Mammalian Genotyping Service
http://www.marshfieldclinic.org/mgs/
Contains information regarding the type of polymorphism (biallelic or multiallelic), validation status, allele length differences, amplification primers, position in the chromosome, and allele frequencies in different population groups

dbSNP or EntrezSNP
http://www.ncbi.nlm.nih.gov/projects/SNP/ or http://www.ncbi.nlm.nih.gov/snp
dbSNP is the NCBI database of genetic variation. It is now incorporated into NCBI's Entrez system and can be queried using the same approach as the other Entrez databases such as PubMed and GenBank. dbSNP also includes Indels (or DIPs) and integrates data from various submitters. Contains general information about the polymorphism, similarly to the Marshfield database, but with extra features and search possibilities. Using custom queries or the "Limits" tab allows restricting your search by chromosome, SNP class, and other criteria (such as heterozygosity, etc.), for example, to select markers with forensic relevant characteristics

UCSC Genome Browser's "Get DNA in Window" tool
http://genome.ucsc.edu/cgi-bin/hgGateway
The "Get DNA in Window" tool included in UCSC Genome Browser allows obtaining a DNA sequence provided the chromosome coordinates. From inside of your target assembly, set the region and submit. Then click on "DNA" in the top blue navigation bar. Of interest is the possibility of using extended case/color options to highlight relevant annotations in that sequence, for example, all SNPs included in dbSNP. This will allow knowing polymorphic positions to avoid during primer design

Primer 3
http://frodo.wi.mit.edu/primer3/

BatchPrimer3
 http://probes.pw.usda.gov/batchprimer3/index.html
Primer 3 is a primer design program available online. It allows designing primers for one input sequence following the parameters indicated by the user. If the sequence includes lowercase letters and the option Lowercase masking is checked, candidate primers having lowercase letter exactly at 3′ end are rejected
BatchPrimer3 is a comprehensive web primer design program to develop different types of primers in a high-throughput manner. It uses the same algorithm as Primer 3 and has the advantage of allowing primer design for a batch of sequences. This can be useful when designing primers for a high number of markers

Table 1
(continued)

Autodimer
http://www.cstl.nist.gov/biotech/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm
Autodimer is a screening tool for primer-dimer and hairpin structures. The program evaluates the
   possible occurrence of secondary structures between the primers included in the input file. It gives
   a schematic representation, Tm and ΔG predictions for the possible interactions. It is of particular
   interest when designing large multiplexes to avoid strong interactions between any primers and
   assure a successful amplification of all markers

BLAST
http://www.ncbi.nlm.nih.gov/blast/

BLAT
 http://genome.ucsc.edu/cgi-bin/hgBlat
BLAST and BLAT are online tools to perform homology searches against sequences deposited in the
   databases. They allow querying PCR primers to know if other similar sequences are found in the
   genome, indicating possible mispriming sites that could reduce the efficiency of the PCR amplification

UCSC Genome Browser's In-Silico PCR tool
http://genome.ucsc.edu/cgi-bin/hgPcr
In-Silico PCR tool, included in UCSC Genome Browser, searches a sequence database with a pair of
   PCR primers using an indexing strategy for fast performance. It allows choosing a template genome
   sequence and test in-silico for the PCR products which would be obtained with a given pair of
   primers. In summary, it permits a quick check for the possible occurrence of unspecific products in
   the PCR, considering a maximum product size. The default maximum product size is 4,000 bp, but
   it can be defined by the user

2. Unlabeled and dye-labeled primers. In order to increase the
   multiplexing level of the technique, a five dye chemistry was
   used. In this particular case, the forward primers were labeled
   using one of the fluorochromes 6-FAM™ (blue), VIC® (green),
   NED™ (yellow), or PET® (red) (Applied Biosystems).
   Delivered lyophilized primers were brought to 100 μM using
   molecular grade water and stored in the dark at –20°C.
   Information on primer sequences is available upon request.

3. Primer mix working solution containing all primers at 1 μM,
   except for markers rs2308137 (2 μM) and rs3047269 (3 μM).
   A stock solution for a large number of reactions can be prepared
   and then divided in smaller aliquots (of volume adequate to
   process the number of reactions desired per PCR) which are
   stored in the dark at –20°C (see Note 2).

4. QIAGEN Multiplex PCR Kit (Qiagen).

5. Hi-Di™ Formamide (Applied Biosystems).

6. GeneScan™ 500 LIZ® Size Standard (Applied Biosystems).

Equipment required:

– Capillary Electrophoresis instrument supporting five dye chemistry (e.g., ABI 310, 3100, or 3130 Genetic Analyzers; Applied Biosystems).

– Fragment analysis software (e.g., GeneMapper ID v3.2 or GeneMapper v4.0; Applied Biosystems).

To prevent cross-contaminations, it is recommended that one should use disposable gloves and aerosol-resistant pipette tips. Pre- and post-PCR samples, reagents, and instruments should be used and stored separately. Formamide is potentially hazard (irritant and teratogen) and should be handled accordingly.

3. Methods

The method described was previously optimized by Pereira et al. (17). Here, it is presented an overview of the stages involved in the development of the Indel multiplex and full details on the genotyping protocol.

3.1. Selecting Indels for Human Identification Purposes

The Indel markers were chosen from a pool of ~4,000 biallelic Indels initially characterized by Weber et al. (7) and included in the Marshfield Diallelic Insertion/Deletion Polymorphisms database website (http://www.marshfieldclinic.org/mgs/). Having in mind their application in genetic identification studies in forensic context, the markers were selected to fulfill the following criteria:

– Autosomal, biallelic Indels.

– Minimum Allele Frequency ≥0.25 in European, African, and Asian population groups.

– Average Heterozygosity ≥0.40.

– Allele length variation of 2–5 bp.

– Localization in intergenic regions (see Note 3).

In summary, the aim was finding noncoding biallellic Indels amenable to analysis in short amplicons, with an expected high degree of polymorphism in most human population groups and widely spread throughout the genome.

3.2. Constructing the Multiplex Assay

The stages involved in constructing the Indel multiplex are summarized in Fig. 2 (see ref. (17) for details). Here the workflow is reviewed highlighting relevant aspects to consider when constructing large Indel multiplexes for forensic applications (see Table 1 for
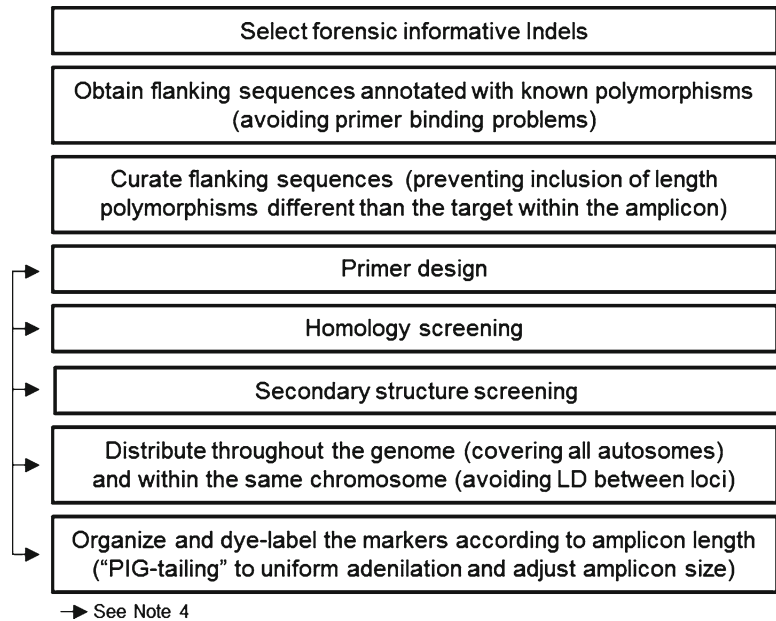
Fig. 2. Schematic representation of the different stages involved in the development of Indel multiplexes: from marker selection to PCR multiplex setup.

details on online databases and computer programs used; on the topic "constructing multiplex assays" see also ref. (20)).

The flanking sequences (±150 bp) of the preselected Indels, including annotations of sequence variants within that region, were obtained using the "Get DNA in Window" tool included in UCSC Genome Browser (human Mar. 2006 assembly; table snp128). This template was used for primer design, and knowing polymorphic positions helped avoiding primer-binding problems and potential null alleles. Furthermore, flanking sequences showing length polymorphisms other than the target, as well as mononucleotide tracts ($\geq$7 bp), were curated in order to prevent its inclusion in the amplicon range and get unexpected length variations.

Primers were designed with Primer3 using parameters aiming short amplicon sizes (60–160 bp), optimum Tm around 60°C, and 45% as minimum CG content. After this, all primer pairs were checked for nonspecific hybridizations in other genome regions using NCBI's BLAST, and markers with primers presenting 100% homology in unexpected locations were excluded.

At this stage, it was necessary to multiplex as many Indels as possible, having simultaneously in mind several aspects of forensic interest:

– Secondary structures screening between primers, allowing an efficient coamplification.

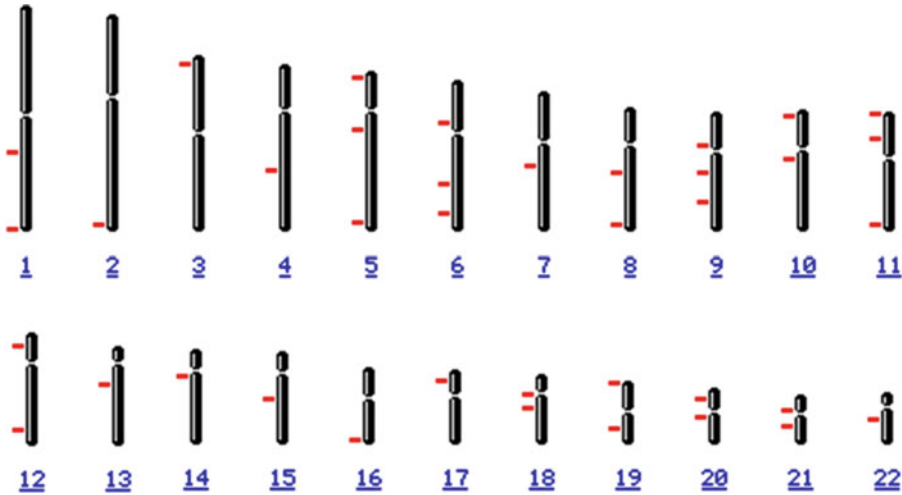– Balanced distribution throughout the human genome, covering all autosomes.

Fig. 3. Distribution of the 38 Indel polymorphisms along the 22 human autosomes.

- Position along the chromosomes, avoiding Linkage disequilibrium between markers at the same chromosome.
- Multicolor dye-labeling and distribution of the markers according to the expected amplicon sizes so that they can be typed in a single electrophoretic run.

At the end of the process, 38 Indels distributed across the 22 human autosomes (see Fig. 3) were successfully multiplexed in a single genotyping reaction. All markers were organized by the expected amplicon length and assigned to four different dye-labeling fluorochromes, enabling a genotyping analysis via a single PCR and electrophoresis. When necessary, tails of random nucleotides were added to the designed primers, aiming to adjust the final amplicon size (see Note 5).

3.3. Loci Information   The 38 Indels included in the assay are presented in Table 2.

3.4. Multiplex PCR Amplification   The amplification of the 38 Indels is performed in a single multiplex PCR using 1× Qiagen Multiplex PCR kit (Qiagen), 0.1 μM of all primers, except rs2308137 (0.2 μM) and rs3047269 (0.3 μM), and 0.3–5 ng of genomic DNA in a 10 μL final reaction volume (17). In Table 3, volumes are indicated for all PCR reagents per reaction, according to the concentration of stock solutions. If a different final volume is used, adjust all reagents accordingly.

The use of a negative control and a positive control sample with known genotype is recommended in every PCR. As reference, the genotypes of 9947A and 9948 human cell lines DNA samples are presented in Table 4.

Thermal cycling conditions are: initial incubation at 95°C for 15 min; followed by 10 cycles at 94°C for 30 s, 60°C for 90 s, and

Table 2
Indel polymorphisms included in the multiplex

Int. code	MID	rs Number	Chromosome	Position (bp) Build 36.3	Build 37.1	Alleles	Amplicon size (S–L)
B1	MID-2719	rs34541393	20	30165066	30701405	–/AACT	57–61
B2	MID-185	rs16624	2	234681130	235016391	–/GT	65–67
B3	MID-1493	rs2307689	19	48896180	44204340	–/TTC	74–77
B4	MID-2946	rs35769550	8	76681235	76518680	–/TGAC	89–93
B5	MID-1504	rs2307700	22	25120901	26790901	–/TCAC	101–105
B6	MID-520	rs140809	10	6027167	5987163	–/CAA	115–118
B7	MID-2305	rs3047269	1	161077452	162810828	–/CTGA	126–130
B8	MID-3221	rs33972805	11	125794082	126288872	–/CT	135–137
B9	MID-2698	rs33917182	20	11643625	11695625	–/CA	142–144
B10	MID-116	rs16402	9	38396788	38406788	–/TTAT	150–154
G1	MID-785	rs1610871	5	171020572	171087970	–/TAGG	61–65
G2	MID-1151	rs2067238	12	113772931	115288548	–/GCT	71–74
G3	MID-1209	rs2067294	9	70504241	71314421	–/CTT	80–83
G4	MID-1514	rs2307710	6	47929222	47821263	–/AGGA	92–96
G5	MID-2050	rs2308242	3	8591709	8616709	–/CT	106–108
G6	MID-1384	rs2307580	9	104626014	105586193	–/AATT	120–124
G7	MID-743	rs1160956	5	65414216	65378460	–/AGA	128–131
G8	MID-3097	rs34511541	18	34677042	36423040	–/CTCTT	143–148
G9	MID-1782	rs2307978	7	83121850	83283913	–/GA	156–158
Y1	MID-2648	rs3051300	17	10076666	10135941	–/GTAT	63–67
Y2	MID-2890	rs10629077	21	30294208	31372337	–/AT	74–76
Y3	MID-3277	rs10688868	11	258180	268180	–/CT	81–83
Y4	MID-1120	rs2067208	16	83139788	84582287	–/GCCAG	93–98
Y5	MID-1383	rs2307579	1	245878706	247812083	–/ATG	104–107
Y6	MID-1824	rs2308020	15	51268809	53481517	–/TT	127–129
Y7	MID-3114	rs3080855	18	21507205	23253207	–/AATT	133–137
Y8	MID-834	rs1610919	12	14801263	14909996	–/AT	142–144
Y9	MID-1643	rs2307839	6	117200251	117093558	–/GA	152–154
R1	MID-1945	rs2308137	6	149655891	149614198	–/GA	61–63
R2	MID-2592	rs36040336	19	1353662	1402662	–/AT	65–67
R3	MID-649	rs1160886	10	54112392	54442386	–/ACT	75–78
R4	MID-1830	rs2308026	4	119404855	119185407	–/CA	83–85
R5	MID-1330	rs2307526	5	5178112	5125112	–/ACAC	95–99
R6	MID-3220	rs34811743	11	30134266	30177690	–/TG	108–110
R7	MID-1997	rs2308189	14	28106508	29036757	–/AACTA	119–124
R8	MID-3031	rs5895447	8	138489776	138420594	–/CA	128–130
R9	MID-1979	rs2308171	13	43778155	44880155	–/TCTG	135–139
R10	MID-2806	rs35605984	21	14556736	15634865	–/TAAAG	151–156

Table 3
PCR amplification mix

Water	3 µL
Qiagen Multiplex PCR master mix (2×)	5 µL
Primer mix (10×)	1 µL
Template DNA 0.3–5 ng/µL	1 µL
Total reaction volume	10 µL

Note: To prevent contamination, it is recommended to use gloves and pipette tips with filter

Table 4
Genetic profiles for the reference samples 9947A and 9948

DNA marker	9947A									
	1	2	3	4	5	6	7	8	9	10
B	SS	SL	SL	SL	SL	LL	SL	LL	SL	SL
G	SL	SL	SL	LL	SL	SL	SS	SL	LL	
Y	LL	LL	SL	SL	SL	SL	SL	SL	SS	
R	SS	SL	SL	LL	LL	SS	SL	SL	SL	SL

DNA marker	9948									
	1	2	3	4	5	6	7	8	9	10
B	SL	SS	SS	SS	SL	SL	SL	SL	SL	LL
G	LL	SL	LL	SL	LL	SS	SS	SL	LL	
Y	SL	SS	LL	LL	LL	SL	SL	SL	SL	
R	SL	SS	LL	SL	SL	SS	SL	SS	LL	SL

S Short allele; *L* long allele

72°C for 60 s; plus 20 cycles at 94°C for 30 s, 58°C for 90 s, and 72°C for 60 s (see Note 6); and a final extension at 72°C for 60 min (see Note 7).

3.5. Analysis of Amplified Products

PCR products are prepared for capillary electrophoresis by adding 1 µL of each amplified product to 10 µL Hi-Di™ Formamide (Applied Biosystems) and 0.3 µL of internal size standard GeneScan™ 500 LIZ® (Applied Biosystems).

Separation and detection are performed by a capillary electrophoresis system. During the optimization of the present method, an Applied Biosystems 3130 Genetic Analyzer was used, with G5 filter set and Performance Optimized Polymer 7 (POP-7™; Applied Biosystems). Analysis conditions were: Module: "FragmentAnalysis 36_POP7"; Oven temperature: 60°C; Injection Voltage: 1.2 kV; Injection Time: 16 s; Run Voltage: 15 kV; Run time: 1,200 s. Data collected are analyzed using a multicomponent matrix (see Note 8).

Finally, the electropherograms are visualized with a fragment analysis software and fragment sizes are determined in order to genotype the samples. In this case, GeneMapper ID v3.2 or GeneMapper v4.0 (Applied Biosystems) were used. It is recommended to create genotyping panels including all markers with the respective allelic bin sets in order to automate allele calling and facilitate the genotyping of a large number of samples. The genotype data can be easily exported to text documents for further treatment and databasing.

An example electropherogram is shown in Fig. 4.
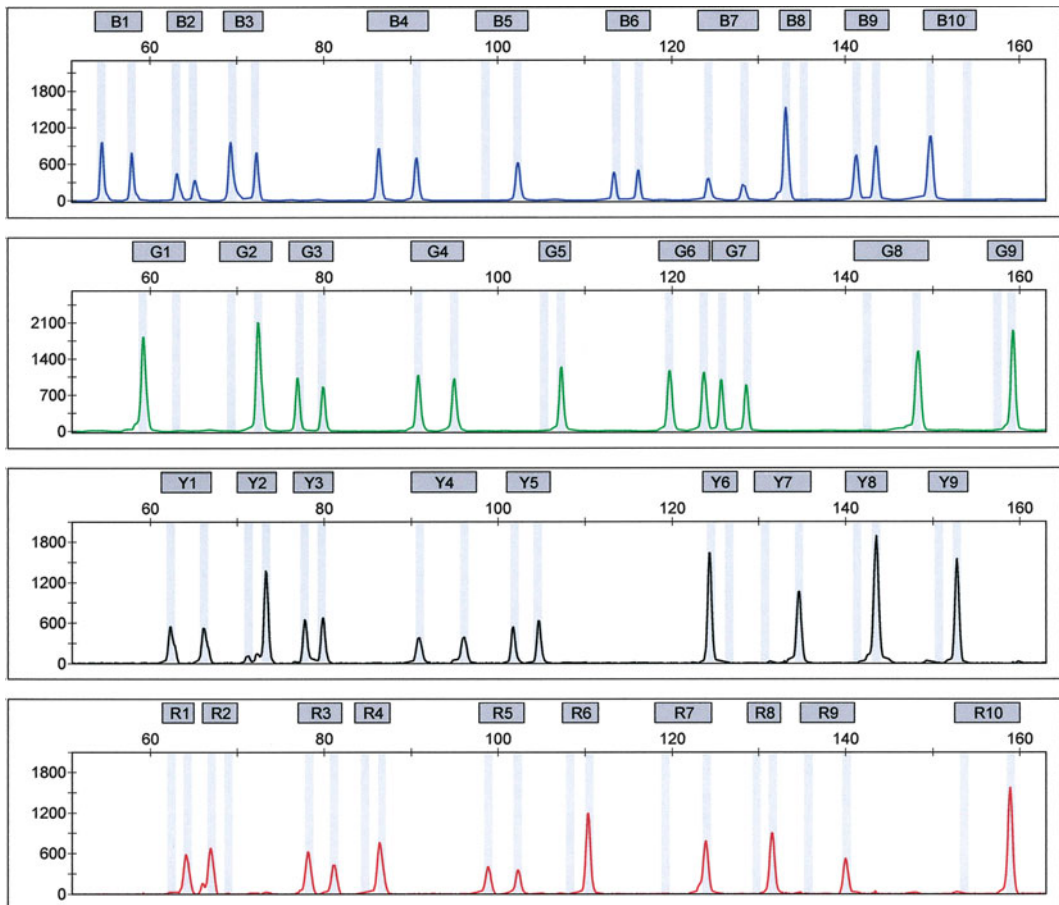
Fig. 4. Example of an electropherogram of the 38 Indel multiplex.

3.6. Additional Considerations

Depending on the particular characteristics of a case, there are important aspects to consider when selecting the best analytical approach, namely the amount of sample available for analysis, the DNA quality (good quality or highly degraded?), and/or the presence of inhibitors. Some recommendations/tips are suggested that might help improving the results, particularly when dealing with challenging DNA:

– In cases of very low DNA quantity or when analyzing highly degraded samples (low template DNA), test increasing the total number of PCR cycles to 32 and/or increasing the primer mix concentration to 1.5 or 2× in the PCR reaction (see Note 9).

– Using larger final volumes in the PCR reaction (e.g., 25 μL) may also improve the results when analyzing highly degraded samples.

– In the presence of PCR inhibition, if enough DNA is available use dilution series of the original sample (e.g., 1:2, 1:5, 1:10; 1:20; 1:50, 1:100). While diminishing the quantity of template DNA in the reaction, it also dilutes possible inhibitors, which is sometimes the only way to obtain a successful amplification.

Occasionally, dye artifacts ("dye-blobs") can appear in electropherograms. Such artifacts are usually due to a poor quality of the primers, produced by an incomplete fluorochrome attaching during manufacturing (20). Also, the primers can suffer degradation with time and freeze–thaw cycles, leading to a reduced performance and the presence of more residual dye artifacts. Minor "dye-blobs" are frequently observed in electropherograms, but usually do not interfere with genotyping. Nevertheless, sometimes they can appear in regions with possible alleles thus hindering genotyping. In these cases, a complementary purification step prior to electrophoresis (e.g., with Sephadex™ G-50 DNA Grade; GE Healthcare) can improve considerably the overall electropherogram by eliminating exceeding primers and dye residues.

If electropherograms present unusually strong peaks, with pull-ups that can be confused with alleles and/or interfere with the internal size standard alignment, testing one of the following is recommended: further dilute the DNA samples; reduce the number of PCR thermal cycles (see Note 6); and apply less PCR product or use less injection time in the capillary electrophoresis system. Good results are usually obtained within the range of 1,000–5,000 rfu, in average.

4. Notes

1. The simultaneous use of STRs and Indels in calculating PI must consider the LD status between the markers under study. Therefore, at least for markers at the same chromosome, an LD test must be performed in the reference population to

discard possible associations between markers (and therefore safely use the product rule in PI calculations); otherwise, linked loci pairs must be treated as haplotypes.

2. Dye-labeled primers are light-sensitive and frequent freeze–thaw cycles can accelerate breakdown of the dye attachment to the oligonucleotide and cause an overall deterioration of the primers leading to weaker performance with time.

3. The noncoding Indels localized only in intergenics regions cannot be found directly in dbSNP. In order to do that, first query the Entrez SNP database for all human Indels fulfilling the criteria: previously validated by assessing allele frequencies, heterozygosity values between 0.40 and 0.50, mapping only once in the human genome, and submitted to the database under the MARSHFIELD handle. Afterwards, obtain the intragenic Indels (obtained by checking-in all *Function Class* search field limits "coding nonsynonymous, intron, coding synonymous, locus region, mrna utr, splice site"). Finally, go to the "History" tab and subtract the results of the first search minus the second (#1 NOT #2). This delivers a list of Indels located only in intergenic regions. To know more about searching for markers of forensic interest in Online databases, see ref. (21).

4. The final stages in Fig. 2 (signaled with arrows) are interconnected and the process runs somehow simultaneously: markers can be excluded for different reasons at different stages and there may be the possibility to rescue an old marker because its prior excluding reason no longer applies (e.g., marker A has conflicting primers with markers B, C, and D and is preliminarily excluded to keep these three; markers B, C, D, E are closely located in a chromosome and only one can be selected to avoid LD; Marker E is selected from the cluster, so Marker A can be rescued). The final aim is obtaining the maximum markers in a single reaction and fulfilling all criteria.

5. More recently, we started to employ the Brownstein et al. (22) "PIG tailing" approach which was originally developed to overcome genotyping problems created by the nonspecific adenylation of amplified products by DNA polymerases, which generates confusing double peaks appearing in the electropherogram (–A,+A). This strategy can be used aiming a dual function of adjusting amplicon sizes while at the same time homogenizing nontemplate adenylation of all products in order to obtain a uniform electropherogram. The unlabeled primers of amplicons that do not need mobility adjustments remain unchanged, or a G is added at the 5′ end if there is not already a G in that position. In the cases requiring little mobility adjustments, a 5′ tail GTTT or partial tails of the original sequences GTTTCTT or GTGTCTT are attached to the unlabeled primer.

6. The number of thermal cycles can be adjusted for different input DNA, depending on the quality and/or concentration of the samples being processed. Samples with a good quantity/quality DNA can provide well-balanced results with a total of 27 cycles, avoiding unusually strong peaks and possible pull-ups, and minimizing adenylation issues. On the contrary, when dealing with very low amounts of DNA (e.g., low copy number DNA) or highly degraded samples, the reaction can benefit of using up to 32 cycles.

7. Double peaks differing in a single bp (–A,+A) can appear due to an incomplete adenylation, usually when high amounts of DNA are present in the PCR reaction. In these cases, apart from reducing the amount of DNA, the final extension step may be extended to 80 min in order to complete adenylation of the amplified products.

8. DS-33 (Dye Set G5) Matrix Standard (Applied Biosystems) is used to generate the multicomponent matrix required when analyzing 6-FAM™, VIC®, NED™, PET®, and LIZ® dye-labeled DNA (for details see the Applied Biosystems 3130 Genetic Analyzer User's Manual).

9. Regularly, concentrations of 0.1 µM are used for all primers in the PCR reaction (except for the two primer pairs needing reinforcement rs2308137 and rs3047269) following the Qiagen Multiplex PCR kit User's Manual, which suggests that for multiplex amplification reactions with more than ten PCR products, a decrease of primer concentration to 0.1 µM may lead to a more uniform product yield (Appendix F: Optimization of Reaction Conditions for Special Multiplex PCR Applications; Large number of PCR products). Nonetheless, the 0.2 µM concentration of each primer, originally suggested at the User's Manual, can be used in cases where the template DNA is scarce or highly degraded, since having more primer available enhances the probability of molecules to encounter and anneal to attain amplification.

Acknowledgments

References

1. Dixon LA, Murray CM, Archer EJ, Dobbins AE, Koumi P, and Gill P. (2005) Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes, *Forensic Sci. Int.* **154**, 62–77.

2. Sobrino B, and Carracedo A. (2005) SNP typing in forensic genetics: a review, *Methods Mol. Biol.* **297**, 107–126.

3. Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, and Morling N. (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification, *Electrophoresis 27*, 1713–1724.

4. Borsting C, Sanchez JJ, Hansen HE, Hansen AJ, Bruun HQ, and Morling N. (2008) Performance of the SNPforID 52 SNP-plex assay in paternity testing, *Forensic Sci Int Genet* **2**, 292–300.

5. Phillips C, Fondevila M, Garcia-Magarinos M, Rodriguez A, Salas A, Carracedo A, and Lareu MV. (2008) Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers, *Forensic Sci Int Genet 2*, 198–204.

6. Krjutskov K, Viltrop T, Palta P, Metspalu E, Tamm E, Suvi S, Sak K, Merilo A, Sork H, Teek R, Nikopensius T, Kivisild T, and Metspalu A. (2009) Evaluation of the 124-plex SNP typing microarray for forensic testing, *Forensic Sci Int Genet 4*, 43-48.

7. Weber JL, David D, Heil J, Fan Y, Zhao C, and Marth G. (2002) Human diallelic insertion/ deletion polymorphisms, *Am. J. Hum. Genet.* **71**, 854–862.

8. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, and Devine SE. (2006) An initial map of insertion and deletion (INDEL) variation in the human genome, *Genome Res. 16*, 1182–1190.

9. Yang N, Li H, Criswell LA, Gregersen PK, Alarcon-Riquelme ME, Kittles R, Shigeta R, Silva G, Patel PI, Belmont JW, and Seldin MF. (2005) Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine, *Hum. Genet. 118*, 382–392.

10. Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, and Feldman MW. (2005) Clines, clusters, and the effect of study design on the inference of human population structure, *PLoS Genet 1*, e70.

11. Bastos-Rodrigues L, Pimenta JR, and Pena SD. (2006) The genetic structure of human populations studied through short insertion-deletion polymorphisms, *Ann. Hum. Genet. 70*, 658–665.

12. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, and Williams SM. (2009) The genetic structure and history of Africans and African Americans, *Science 324*, 1035–1044.

13. Santos NP, Ribeiro-Rodrigues EM, Ribeiro-Dos-Santos AK, Pereira R, Gusmão L, Amorim A, Guerreiro JF, Zago MA, Matte C, Hutz MH, and Santos SE. (2010) Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INDEL) ancestry-informative marker (AIM) panel, *Hum. Mutat. 31*, 184–190.

14. Ribeiro-Rodrigues EM, dos Santos NP, dos Santos AK, Pereira R, Amorim A, Gusmão L, Zago MA, and dos Santos SE. (2009) Assessing interethnic admixture using an X-linked insertion-deletion multiplex, *Am J Hum Biol 21*, 707–709.

15. Edelmann J, Hering S, Augustin C, and Szibor R. (2009) Indel polymorphisms--An additional set of markers on the X-chromosome, *Forensic Science International: Genetics Supplement Series 2*, 510–512.

16. Freitas NS, Resque RL, Ribeiro-Rodrigues EM, Guerreiro JF, Santos NP, Ribeiro-dos-Santos A, and Santos S. (2010) X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel, *Int J Legal Med 124*, 589–593.

17. Pereira R, Phillips C, Alves C, Amorim A, Carracedo A, and Gusmão L. (2009) A new multiplex for human identification using insertion/deletion polymorphisms, *Electrophoresis 30*, 3682–3690.

18. Pimenta JR, and Pena SD. (2010) Efficient human paternity testing with a panel of 40 short insertion-deletion polymorphisms, *Genet. Mol. Res. 9*, 601–607.

19. Pereira R, Phillips C, Alves C, Amorim A, Carracedo Á, and Gusmão L. (2009) Insertion/deletion polymorphisms: A multiplex assay and forensic applications, *Forensic Science International: Genetics Supplement Series 2*, 513–515.

20. Butler JM. (2005) Constructing STR multiplex assays, *Methods Mol. Biol. 297*, 53–66.

21. Phillips C. (2005) Using online databases for developing SNP markers of forensic interest, *Methods Mol. Biol. 297*, 83–106.

22. Brownstein MJ, Carpten JD, and Smith JR. (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping, *Biotechniques 20*, 1004–1006, 1008–1010.

Chapter 11

Capillary Electrophoresis of an 11-Plex mtDNA Coding Region SNP Single Base Extension Assay for Discrimination of the Most Common Caucasian HV1/HV2 Mitotype

Peter M. Vallone

Abstract

The typing of single nucleotide polymorphisms (SNPs) located throughout the human mitochondrial genome assists in resolving individuals with an identical HV1/HV2 haplotype. A set of 11 sites which were selected for distinguishing individuals of a common Western European Caucasian HV1/HV2 mitotype was incorporated into a single base extension (SBE) assay. The assay was optimized for multiplex detection of sequence polymorphisms at positions 3010, 4793, 10211, 5004, 7028, 7202, 16519, 12858, 4580, 477, and 14470 in the mitochondrial genome. PCR primers were designed to allow for multiplex amplification of unique regions in the mitochondrial genome followed by an 11-plex SBE reaction using the SNaPshot® reagent kit. Separation and detection can be accomplished with a capillary-based electrophoresis platform commonly found in most forensic laboratories.

Key words: Single nucleotide polymorphism, PCR, Mitochondrial DNA, Multiplex

1. Introduction

The limitations of mitochondrial DNA's (mtDNA) power of discrimination have been addressed by identifying single nucleotide polymorphisms (SNPs) from the entire mitochondrial genome that provide maximal discrimination among individuals who would otherwise match in HV1/HV2 (1–6). The targeted common HV1/HV2 sequences or "types" comprise ~23% of the Western European Caucasian population. The *most* common Western European Caucasian HV1/HV2 type accounts for ~8% of that population. This most common HV1/HV2 type can be defined as having a G polymorphism at position 263 and a C base

insertion at position 315 (for further details see Table 1 in ref. (2)). Polymorphic sites were identified by sequencing the entire mito-chondrial genome of 241 individuals matching the most common HV1/HV2 types in a population derived from Western European Caucasians (2). From the amount of variation detected, specific panels of SNP sites were proposed as candidates to improve forensic discrimination of specific common HV1/HV2 sequences (2).

An efficient SNP assay protocol for these 11 sites allows practical application in improving the discrimination of mtDNA testing (7). This protocol details the use of a multiplex single base extension (SBE) assay that targets the most useful discriminatory SNP panel identified by Coble et al. (2) for a common Western European Caucasian HV1/HV2 type (263-G, 315.1 ins C). This panel targets 11 sites, 9 from the mtDNA coding region and 2 from the control region outside of HV1/HV2. The SNaPshot® or minisequencing (8) assay consists of a multiplex PCR followed by a multiplex SBE reaction using fluorescently labeled dideoxynucleotide triphosphates (ddNTPs). The fragments are separated and detected by capillary electrophoresis, an instrument common to most forensic laboratories.

2. Materials

2.1. Primers

2.1.1. Ordering of Primers

Oligonucleotides for PCR and SBE assays were purchased from Qiagen Operon (Alameda, CA). Primers were ordered lyophilized, desalted, and unlabeled.

2.1.2. PCR and SBE Primer Sequences

The primer design aspects of the PCR and SBE assays are fully described in Vallone et al. (7). Tables 1 and 2 contain the PCR and SBE primer sequences for the assay, respectively.

2.2. UV Spectrophotomer

A NanoDrop 2000 (UV spectrophotometer) from Thermo Scientific (Wilmington, DE) was used to accurately quantitate primer solutions.

2.3. Thermal Cycler

A GeneAmp® PCR System 9700 (Applied Biosystems, Foster City, CA) was used for thermal cycling and temperature incubation steps. The cycler was run in 9600 emulation mode with a ramp rate of 1°C/s.

2.4. DNA Polymerase and Buffers for PCR

AmpliTaqGold® DNA polymerase, 5 mM Mg^{2+}, and 10× Taq Gold PCR buffer were purchased from Applied Biosystems.

2.5. Deoxyribo-nucleotide Triphosphates

A stock solution containing 10 mM of each deoxyribonucleotide triphosphates (dNTP) was purchased from Promega Corporation (Madison, WI) (see Note 1).

Table 1
PCR primer sequences for multiplex amplification of 11 mitochondrial SNP (mtSNP) sites

Locus	PCR primer sequence	Amplicon size (bp)
477-F 477-R	CTTTTGGCGGTATGCACTTT GGTGTGTGTGTGCTGGGTA	122
3010-F 3010-R	GCGCAATCCTATTCTAGAGTCC TCACGTAGGACTTTAATCGTTGA	124
4580-F 4580-R	TCTTTGCAGGCACACTCATC GCAGCTTCTGTGGAACGAG	130
4793-F 4793-R	CAACCGCATCCATAATCCTT ATGTCAGAGGGGTGCCTTG	186
5004-F 5004-R	TCCATCATAGCAGGCAGTTG TGGTTATGTTAGGGTTGTACGG	124
7028-F 7028-R	GGCCTGACTGGCATTGTATT AAGCCTCCTATGATGGCAAA	125
7202-F 7202-R	ACGCCAAAATCCATTTCACT TTCATGTGGTGTATGCATCG	126
10211-F 10211-R	ACCACAACTCAACGGCTACA GGAGGGCAATTTCTAGATCAAA	143
12858-F 12858-R	ATGATACGCCCGAGCAGA TGTGGGTCTCATGAGTTGGA	126
14470-F 14470-R	CAAGACCTCAACCCCTGACC GGGGGAGGTTATATGGGTTT	129
16519-F 16519-R	ACCACCATCCTCCGTGAAAT AGACCTGTGATCCATCGTGA	183

2.6. Bovine Serum Albumin

Bovine serum albumin (BSA) fraction V lyophilized powder, suitable for molecular biology, non-acetylated (see Note 2) was purchased from Sigma-Aldrich (St. Louis, MO). A stock solution of 3.2 mg/mL was prepared in nanopure water.

2.7. SNaPshot® Reagents

ABI Prism® SNaPshot® Multiplex Kit was purchased from Applied Biosystems. The reagent is a ready-made mix of buffers, fluorescently labeled ddNTPs, and polymerase. The kit is supplied as a 5× concentration master mix.

2.8. Exonuclease I (Exo I) and Shrimp Alkaline Phosphatase

ExoSAP-IT® and shrimp alkaline phosphatase (SAP) were both purchased from USB Corporation (Cleveland, OH) (see Note 3).

2.9. FlashGel®

Amplification of PCR products was confirmed with a FlashGel® system purchased from Lonza (Basel, Switzerland).

Table 2
Single base extension (SBE) primers for the 11 mtSNP sites

Locus	SNP	Primer sequence	[μM][a]
3010-F	G/A	TGTTGGATCAGGACATCCC	0.4
4793-R	A/G	$(T)_4$—TCAGAAGTGAAAGGGGGC	11.5
10211-R	C/T	$(T)_{10}$—ACTAAGAAGAATTTTATGGA	15.5
5004-F	T/C	$(T)_{14}$—A<u>G</u>ACCCAGCTACGCAAAATC[b]	12.4
7028-F	C/T	$(T)_{18}$—GACACGTACTACGTTGTAGC	5.8
7202-F	A/G	$(T)_{22}$—CCACAACACTTTCTCGGCCT	1.0
16519-R	T/C	$(T)_{24}$—TGTGGGCTATTTAGGCTTTATG	5.4
12858-F	C/T	$(T)_{27}$—GCAGCCATTCAAGCAATCCTATA	5.2
4580-R	G/A	$(T)_{29}$—TGGTTAGAACTGGAATAAAAGCTAG	6.0
477-F	T/C	$(T)_{38}$—CCCTCCCACTCCCATACTAC	5.6
14470-R	T/A[c]	$(T)_{41}$—GGGAATGATGGTTGTCTTTGG	10.0

The orientation relative to the revised Cambridge Reference Sequence (rCRS) (9) of each SBE primer is indicated (F or R) after the site name. Poly-T tails of varying (increasing) lengths were added to each SBE primer to enable electrophoretic separation
[a] The concentrations listed are for preparing a *working stock* solution of the 11 SBE primers (see Subheading 3.6)
[b] Underlined base in 5004 is actually an "A" in the rCRS and did not affect the binding of the SBE primer
[c] Note that site 14470 has been observed to also have a C variant

2.10. Capillary Electrophoresis

The ABI PRISM® 3130xl Genetic Analyzer (Applied Biosystems) was used for fragment separation with filter set E5 in order to process the data from the five fluorescent dyes dR110, dR6G, dTAMRA, dROX, and LIZ. An appropriate spectral matrix was created using materials from the matrix standard set DS-02 (Applied Biosystems). A 36-cm array was used for fragment separation (see Note 4).

2.11. Separation Polymer

The performance optimized polymer (POP-6, Applied Biosystems) separation polymer was used for SBE fragment separation (see Note 5).

2.12. Electrophoresis Buffer

A.C.E.™ 10× sequencing buffer purchased from Amresco (Solon, OH) was used in a 1× concentration (diluted with nanopure water).

2.13. Formamide

SBE products were diluted in Hi-Di™ formamide purchased from Applied Biosystems prior to electrophoresis.

2.14. DNA Sizing Standard

GeneScan™ 120 LIZ® Size Standard was purchased from Applied Biosystems (see Note 6).

2.15. Software for SBE Assay Analysis

The commercial software package GeneMapper® HID v3.2 (Applied Biosystems) was used for data analysis (see Note 7).

3. Methods

3.1. Primers

3.1.1. Primer Rehydration

Stock primer solutions (100 μM) were prepared by rehydrating primers with appropriate volumes of a low salt Tris-EDTA buffer (10 mM Tris–HCl and 0.1 mM EDTA) (see Note 8).

3.1.2. Quantitation of Primers

2 μL of a primer stock solution can be quantitated after blanking the instrument with the appropriate buffer (in this case 10 mM Tris–HCl and 0.1 mM EDTA). The quantitation allows for a more accurate measurement of primer concentration and assists in reproducibility of results when new lots of primers are required. The low volume capability of the NanoDrop limits the consumption of the main primer stock solution.

3.2. PCR Amplification Conditions

Multiplex PCR amplification of the 11 unique amplicons was carried out in a total volume of 15 μL (see Note 9). Final PCR reagent concentrations were: 1 unit of AmpliTaqGold® DNA polymerase, 1× Taq Gold PCR buffer, 250 μM dNTPs, 5 mM $MgCl_2$, 0.16 mg/mL BSA fraction V, 0.5 μM of each amplification primer pair (22 total primers).

3.3. Thermal Cycling Conditions

3.3.1. PCR Thermal Cycling Conditions

The PCR thermal cycling conditions represent a "reverse touch-down" approach. Cycling conditions are a slight modification of a protocol previously described for 12-plex amplification reactions (10).

PCR thermal cycling conditions

95°C for 10 min.

   3 Cycles of 95°C for 30 s, 50°C for 55 s, 72°C for 30 s.

   19 Cycles of 95°C for 30 s, 50°C for 55 s +0.2°C per cycle, 72°C for 30 s.

   11 Cycles of 95°C for 30 s, 55°C for 55 s, 72°C for 30 s.

72°C for 7 min.

25°C hold.

3.3.2. SBE Thermal Cycling Conditions

Thermal cycling conditions for SBE reactions were carried out exactly as described in the SNaPshot® multiplex kit manual.

SBE thermal cycling conditions

25 Cycles of 96°C for 10 s, 50°C for 5 s, 60°C for 30 s.

3.4. Confirmation of PCR Products

Confirmation of multiplex PCR products is optional, but the presence of amplicons can be checked on a FlashGel® cassette. The 1.2% agarose, 13-lane cassettes were used according to the manufacturer's instructions. The system can separate 10–4,000 bp DNA fragments in 2–7 min with a self-contained detection system.

Due to the similarity in amplicon sizes, one cannot typically uniquely distinguish between all 11 products. This process acts as a general confirmation of PCR success before continuing with the SBE reaction (see Note 10).

3.5. PCR Product Purification

Following PCR thermal cycling, unincorporated primers and dNTPs were degraded and inactivated by adding 6 µL of ExoSAP-IT® to the 15 µL post-PCR reaction (2 µL of ExoSAP-IT® per 5 µL of post-PCR reaction). Reactions were mixed briefly and incubated at 37°C for 90 min and then 80°C for 20 min. An extended incubation at 37°C was required to ensure digestion of all unincorporated PCR primers (see Note 11).

3.6. SBE Conditions

Multiplex primer extension reactions were carried out in a total volume of 10 µL. Typically in our work with reference samples, a one-half (1/2×) reaction equivalent of the SNaPshot® reagent was used for SBE reactions (see Note 12). SBE reaction components per reaction were as follows: 2.5 µL of SNaPshot® multiplex kit mix, 0.5 µL of the 10× AmpliTaqGold® PCR buffer, 3 µL of multiplex PCR products (see Note 13), 3 µL of deionized water, and 1 µL of the *working stock* solution of SBE primers (see Table 2 for SBE *working stock* primer concentrations).

3.7. SBE Product Purification

Excess fluorescently labeled ddNTPs were inactivated by addition of 1 unit of SAP. Reactions were mixed briefly and incubated at 37°C for 30 min then 80°C for 20 min (see Note 14).

3.8. Preparation of SBE Products for Electrophoretic Separation and Detection

Fluorescently labeled SBE products were prepared for CE analysis by mixing 14 µL of Hi-Di formamide, 0.4 µL of the LIZ-120 internal sizing standard, and 1.0 µL of SAP-treated SBE products.

3.9. Capillary Electrophoresis Conditions

A 36-cm capillary array filled with denaturing POP-6 polymer was utilized for DNA fragment separation. A.C.E. capillary electrophoresis running buffer was used in 1× concentration. The run module parameters were: Run temperature = 60°C, capillary fill volume = 184 steps, prerun voltage = 15 kV, prerun time = 60 s, injection voltage = 1 kV, injection time = 13 s, run voltage = 15 kV, data delay = 200 s, and run time = 1,200 s.

3.10. Data Analysis

Data were imported into GeneMapper HID v3.2. The appropriate size standard (LIZ-120) and custom panel for the mitochondrial
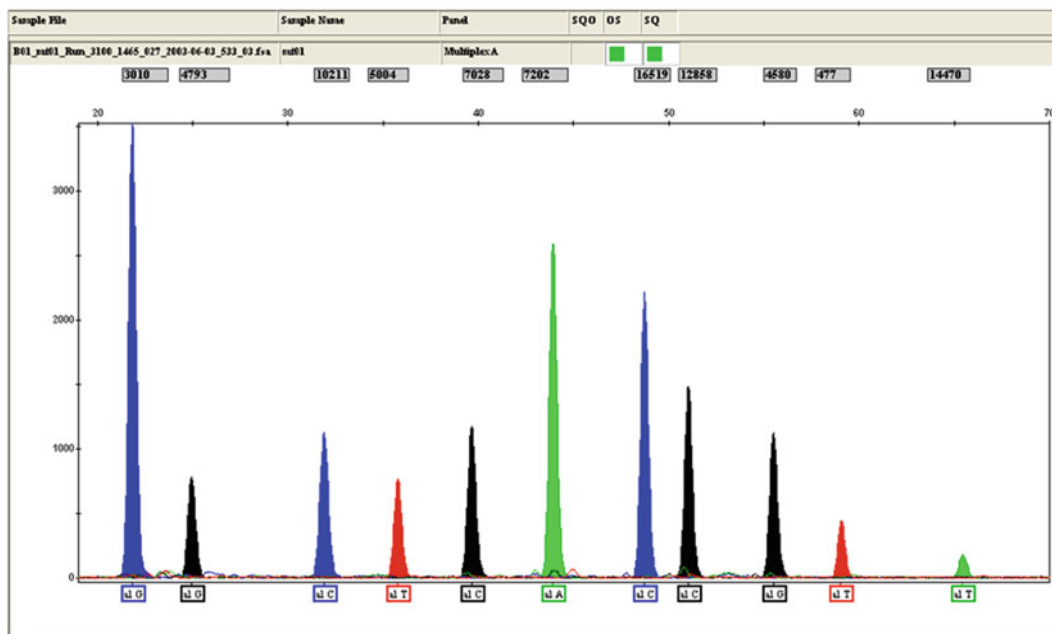
Fig. 1. An example electropherogram for the 11-plex SBE assay. Eleven unique SBE primers were separated and detected. The *numbers* at the top of the plot indicate the position of the polymorphism in the mitochondrial genome (3010, 4793, etc.). The fragment size is defined by the length of the SBE primer sequence (typically between 20 and 70 nucleotides). Below each peak is the base call for that specific locus. The base call is defined by the fluorescent dye detected (*Blue* = G, *Black* = C, *Red* = T, and *Green* = A). Note that for sites 4793, 10211, 16519, 4580, and 14470, the SBE primer binds to the PCR products in the *reverse* orientation. Therefore, the complimentary dideoxynucleotide triphosphates (ddNTPs) were enzymatically added to the reverse SBE primers. Allele calls for the five reverse binding SBE primers were adjusted in GeneMapper.

SNP analysis were applied (see Fig. 1). This resulted in labeling of the peaks of interest with the corresponding allele calls. It is possible to manually determine allele calls by careful visual inspection, but the automated process is less time-consuming and error prone (see Note 15).

4. Notes

1. Once obtained from the manufacturer, the stock solution of dNTPs (in about 1.5 mL) was aliquoted into smaller volumes (e.g., 50 µL) and stored at –20°C. A fresh aliquot was used for each PCR mastermix. This technique reduced the number of freeze–thaw cycles for the main stock and helped preserve the integrity of the dNTPs.

2. It has been suggested in the literature (11) that acetylated BSA has the potential to inhibit the PCR reaction. Highly purified non-acetylated BSA should be used in PCR assays.

3. ExoSAP-IT reagent can also be prepared by combining Exonuclease I (1.4 µL where 1 µL = 10 units) and SAP (2.6 µL where 1 µL = 1 unit). Results can be compared to the commercial ExoSAP-IT reagent to confirm activity.

4. A 22-cm multicapillary array (Applied Biosystems) may be used if the instrument is dedicated to SBE assays and not sequencing or STR fragment separation. The shorter array length results in faster separation times and slightly less polymer usage.

5. POP-4 and POP-7 can also be used. Note that the mobility of the SBE fragments will slightly change depending on the polymer. The bins and panels in the GeneMapper software will also have to be appropriately adjusted to reflect any mobility changes.

6. Allelic ladders are not required for this SBE SNP detection assay. Ideally, a set of control samples containing all expected allelic variations may be used to construct the proper bins and panels for automated analysis.

7. The newer GeneMapper® ID-X (or higher) software does not support the creation of bins and panels for SNP markers. Alternatively, a copy of the older GeneScan/Genotyper software or GeneMapper 4.0 (Applied Biosystems) can be used for SNP analysis.

8. Primer storage in a low salt buffer at 4°C resulted in primer stability over a period of 12–18 months with no changes in assay performance.

9. The PCR reaction volume can be increased to 25 µL if needed. Increased PCR volumes can assist in diluting out effects of PCR inhibitors found in casework samples.

10. Any preferred method of separation and detection of PCR products can be used for this purpose [e.g., Bioanalyzer 2100 (Agilent), agarose gel, etc.]. Do not use UV spectroscopy for the detection of PCR products unless the PCR reaction has been physically purified (microscreen filter, microcon, or centricon). Absorbance interference from dNTPs and PCR primers (degraded or nondegraded) will skew results.

11. It is important that all PCR primers are fully degraded prior to performing the SBE reaction. Remaining PCR primers will compete for SBE reagents and will be extended as SBE primers. Unwanted extension of PCR primers will result in interfering peaks in the fragment range of approximately 20 (±5) nucleotides.

12. A 1× equivalent of the SNaPshot reagent can be used for enhanced sensitivity or if a higher level multiplex (more loci) is being developed.

13. The volume of post-PCR products required can vary slightly depending on the efficiency of the multiplex PCR. Experiments should be performed to empirically determine the optimal volume of PCR products required for input into the SBE reaction.

14. If not fully degraded by the SAP treatment, the unincorporated ddNTPs will migrate along with the SBE products, complicating data interpretation.

15. The paper (7) describing this 11-plex assay in addition to GeneMapper® bins and panel for running the SBE assay with POP-6 can be found online at: http://www.cstl.nist.gov/biotech/strbase/mtDNA.htm).

References

1. Parsons, T.J. and Coble, M.D. (2001). Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. Croat Med J **42**, 304–309.

2. Coble, M.D., Just, R.S., O'Callaghan, J.E., Letmanyi, I.H., Peterson, C.T., Irwin, J.A., Parsons, T.J. (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.

3. Just, R.S., Leney, M.D., Barritt, S.M., Los, C.W., Smith, B.C., Holland, T.D., and Parsons, T.J. (2009). The use of mitochondrial DNA single nucleotide polymorphisms to assist in the resolution of three challenging forensic cases. J Forensic Sci. **54**, 887–91.

4. Coble, M.D., Vallone, P.M., Just, R.S., Diegoli, T.M., Smith, B.C., and Parsons, T.J. (2006). Effective strategies for forensic analysis in the mitochondrial DNA coding region. Int J Legal Med. **120**, 27–32.

5. Allen, M., Andreasson, H. (2005). Mitochondrial D-loop and coding sequence analysis using pyrosequencing. Methods Mol Biol **297**, 179–196.

6. Budowle B, Gyllensten U, Chakraborty R, Allen M (2005) Forensic analysis of the mitochondrial coding region and association to disease. Int J Legal Med **119**, 314–315. Epub Apr 21, 2005.

7. Vallone, P.M., Just, R.S., Coble, M.D., Butler, J.M., Parsons, T.J. (2004). A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. *Int J Legal Med.* **118**, 147–157.

8. Tully, G., Sullivan, K.M., Nixon, P., Stones, R.E., and Gill, P. (1996). Rapid Detection of Mitochondrial Sequence Polymorphisms Using Multiplex Solid-Phase Fluorescent Minisequencing. Genomics **34**, 107–113.

9. Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet **23**, 147.

10. Bell, P.A., Chaturvedi, S., Gelfand, C.A., Huang, C.Y., Kochersperger, M., Kopla, R., Modica, F., Pohl, M., Varde, S., Zhao, R.B., Zhao, X.J., and Boyce-Jacino, M.T. (2002). SNPstream UHT: Ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. Biotechniques Suppl 70–77.

11. McKeown, B.J. (1994). An acetylated (nuclease-free) bovine serum albumin in a PCR buffer inhibits amplification. BioTechniques **17**, 246–248.

Chapter 12

Capillary Electrophoresis of a Multiplex Reverse Transcription-Polymerase Chain Reaction to Target Messenger RNA Markers for Body Fluid Identification

Cordula Haas, Erin Hanson, and Jack Ballantyne

Abstract

The analysis of cell-specific mRNA expression is a promising new method for the identification of body fluids. A number of mRNA markers have been identified for the forensically most relevant body fluids: blood, saliva, semen, vaginal secretions, and menstrual blood. Apart from a significant improvement in specificity compared to conventional protein-based methods, other important advantages of body fluid identification by mRNA profiling include the possibility of simultaneously isolating RNA and DNA from the same piece of stain and the ability to multiplex numerous RNA markers for the identification of one or several body fluids. RNA profiling can be incorporated into current DNA analysis pipelines.

Key words: Forensic science, mRNA profiling, Body fluid identification

1. Introduction

Identification of the tissue source of a biological stain can be an important issue in forensic casework. The presence of biological material such as blood, semen, and saliva stains can indicate the location of potential sources of DNA that, once recovered, is used to identify the donor of the biological material. Often the identification of biological material is crucial to the investigation and prosecution of the case.

Conventional methods for body fluid stain analysis involve the use of enzymatic or immunologic tests performed in a serial (as opposed to parallel) manner. There are additional disadvantages associated with conventional testing to confirm the presence of a particular body fluid including poor timeliness and limited specificity

(i.e., cross-reactivity with other species or tissues). Furthermore, no such tests are available for the definitive identification of vaginal secretions, menstrual blood, urine, and skin cells. Messenger RNA (mRNA) profiling is a promising new technique for the identification of body fluids in biological stains. It is based on the premise that each single tissue type is comprised of cells that have a unique transcriptome or gene expression (i.e., mRNA) profile. A number of markers have been identified for the forensically most relevant body fluids: blood, saliva, semen, vaginal secretions, and menstrual blood (see Table 1) (1–23).

RNA is notorious for its rapid postmortem and in vitro decay (24), but several reports have pointed out an unexpectedly high stability of RNA in forensic stains (25–27). Since casework material is often limited, an important advantage of body fluid identification by mRNA profiling is the possibility of simultaneously isolating RNA and DNA from the same stain (5, 20, 23). A list of successfully applied extraction protocols is shown in Table 2, two of which are described in detail in the Material and Methods sections.

It is beneficial to quantify the RNA and put a defined amount (around 1–10 ng) into the reverse transcription (RT) reaction in order to avoid false positive and negative signals. In our experience, the Nanodrop Spectrophotometer and the Agilent Bioanalyzer are not suitable for assessing the RNA quantity and quality in forensic samples (17). The quantity of the extracted RNA from forensic stains is usually below the detection limit of the Nanodrop spectrophotometer. Although the RNA quality of forensic samples, as assessed with an Agilent Bioanalyzer, is often poor (according to the RIN metric), after reverse transcription and marker-specific amplification, electropherograms nevertheless reliably showed fragments of the expected sizes. Two sensitive and convenient fluorescence-based RNA quantification methods (7, 22) are described in the Material and Methods sections, but might have limited casework utility because they are not human-specific. Thus, bacterial RNA, which is found in abundance in buccal or vaginal swabs, will result in an underestimate of the amount of human RNA in these sample types. Accurate quantification of human total RNA in forensic samples awaits the development of a sensitive and robust human-specific RNA quantification system that employs a consistently expressed housekeeping gene.

For the RT reaction, several kits are available and their suitability for use with forensic biological samples has been demonstrated (see Table 2) (19, 23). One method is described in detail in the Material and Methods section. A RT minus control (no reverse transcriptase added) can be used to detect possible DNA contamination.

Another advantage of mRNA profiling is the ability to multiplex numerous markers for the identification of one or several body fluids. The development of a multiplex PCR can be a challenge, in that not all PCR products are amplified with equal efficiency, and nonspecific

Table 1
Genes that have been reported to be expressed in a tissue-specific manner or body fluid-specific bacteria

Gene	Function/localization of protein	References
Blood		
Hemoglobin β (HBB)	β-Subunit of hemoglobin A	(10, 16, 19, 22, 23)
Hemoglobin α (HBA)	α-Subunit of hemoglobin A	(8, 22, 23)
Aminolevulinate synthase 2 (ALAS2)	Erythroid-specific mitochondrial enzyme that catalyzes the first step in the heme biosynthetic pathway	(11, 19, 22, 23)
CD3 γ molecule (CD3G)	Part of the T-cell receptor-CD3 complex, which plays an important role in coupling antigen recognition to several intracellular signal-transduction pathways	(10, 22, 23)
Ankyrin 1 (ANK1)	Erythrocyte membrane protein, which provides the primary linkage between the membrane skeleton and the plasma membrane	(9, 22, 23)
Erythroid form of porpho-bilinogen deaminase (PBGD)	Third enzyme of the heme biosynthetic pathway	(7, 16, 19, 22, 23, 26)
β-spectrin (SPTB)	Erythrocyte membrane protein	(5, 7, 10, 11, 16, 19, 22, 23, 26)
Glycophorin A (Glyco A)	Glycophorin A is a major sialoglycoprotein of the human erythrocyte membrane and carries the M or N group antigen	(20)
Saliva		
Histatin 3 (HTN3)	Part of the nonimmune host defense system in the oral cavity	(4, 5, 7, 11, 16, 20, 26)
Statherin (STATH)	Salivary protein that binds hydroxyapatite and acts as an inhibitor of precipitation of calcium phosphate salts in the oral cavity	(4, 7, 11, 16, 20, 26)
Mucin 7 (MUC7)	Small salivary mucin, which is thought to play a role in facilitating the clearance of bacteria in the oral cavity and to aid in mastication, speech, and swallowing	(29)
Streptococcus salivarius	Oral bacterial strain	(18)

(continued)

Table 1
(continued)

Gene	Function/localization of protein	References
Semen		
Protamine 1 (PRM1)	During spermatogenesis, histones are replaced by protamines, which are	(3, 5, 7, 11, 16, 17, 26)
Protamine 2 (PRM2)	essential for the highly condensed chromatin structure in the sperm nucleus	(3, 7, 10, 11, 16, 17, 20, 26)
Transglutaminase 4 (TGM4)	The transglutaminase family catalyzes the irreversible cross-linking of peptide-bound glutamine residues either with peptide-bound lysines or with primary amines. TGM4 is specific to the prostate	(9, 20)
Prostate-specific antigen, Kallikrein 3 (PSA)	Serine protease, found in seminal fluid, prostatic fluid, male serum, male and female urine	(8, 10, 17)
Semenogelin 1 (SEMG1)	Protein originating in the seminal vesicles	(9, 10, 17)
Vaginal secretions		
Mucin 4 (MUC4)	Mucins protect the surfaces of the reproductive tract epithelium from pathogen penetrance and modulate sperm entry into the uterus	(5, 7, 8, 16, 26)
Human β-Defensin 1 (HBD1)	Antimicrobial peptide of urogenital tissues	(7, 16)
Lactobacillus crispatus	Lactobacilli are specific to the vagina	(21)
Lactobacillus gasseri		(21)
Menstrual blood		
Matrix Metalloproteinase 7 (MMP7)	Endopeptidases involved in the breakdown of extracellular matrix components	(2, 7, 11, 13, 16)
Matrix Metalloproteinase 10 (MMP10)		(2, 9, 11)
Matrix Metalloproteinase 11 (MMP11)		(2, 6, 13, 16, 20)

Table 2
List of successfully applied RNA extraction and RT methods from two collaborative exercises (19, 23)

RNA extraction	Reverse transcription (RT)
Allprep RNA/DNA mini-Kit (Qiagen)*	SuperScript III (Invitrogen)*
mirVana miRNA Isolation Kit, RNA+DNA (Ambion)	RETROscript Kit (Ambion)
RNeasy mini-Kit (Qiagen)	First-Strand cDNA Synthesis Kit (Novagen)
RNeasy micro-Kit (Qiagen)	MMLV RT (Ambion)
RNAqueous Kit (Ambion)	High capacity cDNA RT Kit (AB)
NucleoSpin RNA XS (Macherey-Nagel)	SuperScript VILO cDNA Synthesis kit (Invitrogen)
PureLink RNA mini-Kit (Invitrogen)	ProtoScript First-Strand cDNA Synthesis Kit (NEB)
Guanidine isothiocyanate-Phenol/Chloroform (manual)*	PrimeScript (Takara)

Methods marked with an *asterisk* (*) are described in detail in the Material and Methods sections

background products are often observed. Development of a robust multiplex system involves the consideration and optimization of numerous factors such as marker abundance (high, medium, low), primer annealing temperatures, amplicon sizes, labeling dyes, buffer composition, cycle number, and amount of input template. In the Material and Methods section, an easy-to-use commercially available multiplex PCR kit is described and an example singleplex protocol is provided.

Post-PCR purification can increase peak heights (depending on the elution volume) and eliminate dye blobs and primer peaks (28). It does not allow for detection of additional markers that were not observed prior to post-PCR purification, but results in increased peak heights and reduced baseline signal noise. Therefore, the use of post-PCR purification is recommended for low template samples.

A positive control consisting of a marker expressed in all tissue types would confirm a successful analysis. However, currently no housekeeping gene has been described that is suitable for the detection of all body fluids. Saliva and semen, for example, show very little or no expression of common housekeeping genes (GAPDH, 18S rRNA, TEF, UCE), presumably because of limited cell metabolism in spermatozoids and in the desquamated cells of the cheek mucosa (16, 20). Additionally, the high abundance of housekeeping genes relative to mRNA may result in the titration of critical PCR components, thereby affecting the overall sensitivity of the multiplex system. Coextracted DNA might serve as a positive control for the presence of human cells. A concurrent negative

RNA result would suggest that the tested body fluid is absent or that any body fluid-specific mRNA is degraded and is undetectable. In the case of degradation, housekeeping genes would also be affected.

2. Materials

2.1. General Supplies and Equipment

1. RNaseZap (Ambion/Applied Biosystems AB).
2. Spin baskets (e.g., Promega).
3. Microcentrifuge tubes (e.g., DNA LoBind tubes, Vaudaux-Eppendorf, see Note 2).
4. Tweezers, scissors, and/or scalpels (single use or cleaned thoroughly between handling of different samples).

2.2. Manual RNA Extraction

1. Denaturing solution: 4 M guanidine thiocyanate, 0.02 M sodium citrate dihydrate, 0.5% sarkosyl, in nuclease-free water.
2. 2 M Sodium acetate solution, pH 4.0, in nuclease-free water.
3. Wash solution: 75% ethanol/25% DEPC-treated water.
4. Acid phenol chloroform (Ambion).
5. GlycoBlue carrier (Ambion).
6. Isopropanol.
7. RNAsecure Resuspension Solution (Ambion).
8. DNase treatment: TURBO DNA-free kit (Ambion).

2.3. DNA/RNA Coextraction with the AllPrep DNA/RNA Mini-Kit (Qiagen)

1. Buffers RPE, AW1, and AW2 (supplied as concentrate; replenished with ethanol before use according to manufacturer's instructions).
2. β-Mercaptoethanol (added to Buffer RLT Plus before use: 10 μL β-mercaptoethanol per 1 mL Buffer RLT Plus; stable for 1 month at room temperature).
3. Optional (see Note 1): Carrier-RNA working solution (from Qiagen RNA micro-Kit): dilute stock solution (310 ng/μL) to a final concentration of 4 ng/μL in RLT Plus-Buffer.
4. On-column DNase digestion using the RNase-free DNase Set (Qiagen): To prepare DNase I stock solution, add 550 μL RNase-free water to the vial using an RNase-free needle and syringe, mix gently by inverting the vial (do not vortex!). Aliquots can be stored for up to 9 months at –20°C.

2.4. RNA Quantification, Reverse Transcription, and PCR

1. Quant-iT RiboGreen RNA Kit (Invitrogen), to use with a fluorescence microplate reader, low-range assay 1–50 ng/mL, high-range assay 20–1,000 ng/mL.
2. Quant-iT RNA Assay (Invitrogen), to use with the Qubit fluorometer (Invitrogen), assay range 20–500 ng/mL.

3. Superscript III First-Strand Synthesis System (Invitrogen).

4. Multiplex PCR kit (Qiagen).

5. Singleplex reagents: 10× PCR Buffer I, dNTPs (2.5 mM each), AmpliTaq Gold DNA-Polymerase 5 U/µL (all from AB) (see Table 3).

2.5. Post-PCR Purification and Capillary Electrophoresis

1. MinElute PCR Purification Kit (Qiagen).

2. Hi-Di Formamide (AB).

3. GeneScan-500 LIZ or ROX Size Standard (AB).

4. Performance-optimized polymers POP-4, POP-6, POP-7 (AB).

3. Methods

Special precautions have to be taken to avoid the degradation of RNA by the ubiquitously present RNases, namely a separate working place only for RNA, separate pipettes, RNase-free plastic ware, cleaning of the whole working area with RNaseZap, and the regular changing of gloves.

3.1. Manual RNA Extraction

1. Preheat 500 µL denaturing solution + 3.6 µL β-mercaptoethanol per sample at 56°C for 10 min.

2. Cut a piece of stain (on swab/textile) with a sterile scalpel/ scissors/tweezers and transfer to a 1.5 mL LoBind Tube (see Note 2).

3. Add 503 µL solution to each tube, vortex, and incubate at 56°C for 30 min.

4. Transfer swab/textile with clean tweezers into spin basket and centrifuge 5 min at 13,000 rpm. Discard the spin basket including the swab/textile (see Note 3).

5. Add 50 µL 2 M sodium acetate and 600 µL acid phenol chloroform to each sample, invert to mix, and incubate at 4°C for 30 min to 1 h. Centrifuge for 20 min at 13,000 rpm.

6. Pipet RNA-containing aqueous phase (upper phase, see Note 4) into new 1.5 mL LoBind tube. Discard bottom layer.

7. Add 2 µL GlycoBlue carrier (see Note 5) and 500 µL isopropanol to aqueous layer, invert to mix, precipitate at –20°C for at least 1 h (or overnight).

8. Centrifuge for 20 min at 13,000 rpm. Remove and discard supernatant.

9. Wash pellet with 900 µL wash solution, centrifuge for 10 min at 13,000 rpm, then remove supernatant. Dry sample in vacuum centrifuge for 3–5 min or at room temperature for several hours (in fume hood or otherwise protected from environment).

Table 3
Primers for the amplification of gene-specific sequences

Gene	Primer sequence	Size (bp)	References
Blood			
HBB	f: GCA CGT GGA TCC TGA GAA C	61	(16)
	r: ATG GGC CAG CAC ACA GAC		
HBA	f: ACG CTG GCG AGT ATG GT	112	(22)
	r: CCC TTA ACC TGG GCA GAG		
ALAS2	f: TGT GTC CGT CTG GTG TAG TA	133	(19)
	r: AAA CTT ACT GGT GCC TGA GA		
CD3G	f: GTC GAG AGC TTC AGA CAA GC	154	(22)
	r: AGG AGG AGA ACA CCT GGA CT		
ANK1	f: GGC ATG CCC TAT TCT GTG	165	(22)
	r: CTT AGA AGC CAG ATG CAA GC		
PBGD	f. TGG ATC CCT GAG GAG GGC AGA AG	177	(7)
	r: TCT TGT CCC CTG TGG TGG ACA TAG CAA T		
SPTB	f: AGG ATG GCT TGG CCT TTA AT	247	(7)
	r: ACT GCC AGC ACC TTC ATC TT		
Glyco A	f: CAG ACA AAT GAT ACG CAC AAA CG	188	(20)
	r: CCA ATA ACA CCA GCC ATC ACC		
Saliva			
HTN3	f: GCA AAG AGA CAT CAT GGG TA	134	(7)
	r: GCC AGT CAA ACC TCC ATA ATC		
STATH	f: TTT GCC TTC ATC TTG GCT CT	93	(16)
	r: CCC ATA ACC GAA TCT TCC AA		
MUC7	f: CTA AAA GCA AGC AAC TGG AT	197	(29)
	r: AAG TGA GAT TTG GGT GAT TG		
S. salivarius	f: GTG TTG CCA CAT CTT CAC TCG CTT CGG	544	(18)
	r: CGT TGA TGT GCT TGA AAG GGC ACC ATT		
Semen			
PRM1	f: GCC AGG TAC AGA TGC TGT CGC AG	153	(7)
	r: TTA GTG TCT TCT ACA TCT CGG TCT		
PRM2	f: GGC GCA AAA GAC GCT CC	91	(7)
	r: GCC CAG GAA GCT TAG TGC C		
TGM4	f: TGA GAA AGG CCA GGG CG	215	(20)
	r: AAT CGA AGC CTG TCA CAC TGC		
PSA	f: TGT CCG TGA CGT GGA TTG	82	(17)
	r: GGT TGG GAA TGC TTC TCG		
SEMG1	f: TCG GTA ACC ATG TGA AAG GA	120	(17)
	r: GTG TCA TCC ATG GAC CAA GA		
Vaginal secretions			
MUC4	f: GGA CCA CAT TTT ATC AGG AA	235	(7)
	r: TAG AGA AAC AGG GCA TAG GA		
HBD1	f: CCC AGT TCC TGA AAT CCT GA	215	(16)
	r: CAG GTG CCT TGA ATT TTG GT		
L. crispatus	f: CAG AGC AAG CGG AAG CAC A	279	(21)
	r: CAT CTC TGC ATT GGG TTC CC		
L. gasseri	f: GAG AAA GCC AAG CGG AAG C	253	(21)
	r: TTG CTT ACT TAC TGC TCC CCG		

Table 3
(continued)

Gene	Primer sequence	Size (bp)	References
Menstrual blood			
MMP7	f: TCA ACC ATA GGT CCA AGA AC	240	(7)
	r: CAA AGA ATT TTT GCA TCT CC		
MMP10	f: ACA GGG AAG CTA GAC ACT GA	227	(29)
	r: CTG GAG AAT GTG AGT GGA GT		
MMP11	f: GGT GCC CTC TGA GAT CGA C	92	(16)
	r: TCA CAG GGT CAA ACT TCC AGT		

Primers are designed to overlap exon-exon-junctions or span an intron in order to ensure that the obtained products are not due to the presence of contaminating DNA

10. Resuspend the RNA pellet in 20 µL nuclease-free water or preheated RNAsecure Resuspension Solution (see Note 6), at 60°C for 10 min.

11. DNase treatment (see Note 7): Add 2 µL 10× Buffer and 1 µL DNase to 20 µL RNA in a 0.5 mL tube and incubate at 37°C for 20–30 min. Add 2.3 µL Inactivation Reagent and mix well. Incubate at room temperature for 5 min, mixing occasionally. Centrifuge for 1.5 min at 10,000 rpm and pipet ~20 µL RNA into new tube.

3.2. DNA/RNA Coextraction with the Qiagen AllPrep DNA/RNA Mini-Kit

Perform all steps of the procedure at room temperature and all centrifugation steps in a standard microcentrifuge at 20–25°C.

1. Cut a piece of stain (on swab/textile) with a sterile scalpel/scissors/tweezers and transfer to a 1.5 mL LoBind Tube (see Note 2).

2. Add 350 µL RLT Plus-Buffer (containing β-ME).

3. Optional: add 5 µL Carrier-RNA working solution (see Note 1).

4. Vortex 30 s and incubate up to 3 h at 56°C (see Note 8).

5. Transfer swab/textile with clean tweezers into spin basket and centrifuge 5 min at 13,000 rpm. Discard the spin basket including the swab/textile (see Note 3).

6. Transfer the lysate into a new AllPrep DNA spin column (purple) and centrifuge 30 s at 11,000 rpm.

7. Place the AllPrep DNA spin column in a new 2 mL collection tube and store at room temperature for later DNA purification steps 16–20. Use the flow-through for RNA purification steps 8–15.

RNA purification:

8. Add 350 µL 70% Ethanol to the flow-through and mix well by pipetting.

9. Transfer all 700 µL to an RNeasy spin column (pink) and centrifuge 15 s at 11,000 rpm. Discard flow-through.

10. Add 350 µL Buffer RW1 to the RNeasy spin column and centrifuge 15 s at 11,000 rpm. Discard flow-through.

11. On-column DNase digestion (see Note 7): per sample, mix 10 µL DNase I stock solution with 70 µL Buffer RDD and add the DNase I incubation mix (80 µL) directly to the RNeasy spin column membrane. Incubate at room temperature for 15 min.

12. Add 350 µL Buffer RW1 to the RNeasy spin column. Centrifuge 15 s at 11,000 rpm. Discard flow-through.

13. Add 500 µL Buffer RPE to the RNeasy spin column. Centrifuge 15 s at 11,000 rpm. Discard flow-through.

14. Add 500 µL Buffer RPE to the RNeasy spin column. Centrifuge 2 min at 11,000 rpm. Discard flow-through and collection tube. Place the RNeasy spin column in a new 2-mL collection tube and centrifuge 1 min at full speed.

15. Place the RNeasy spin column in a 1.5-mL collection tube. Add 30 µL RNase-free water directly to the RNeasy spin column membrane. Centrifuge 1 min at 11,000 rpm to elute the RNA. Discard the column.

DNA purification:

16. Add 500 µL Buffer AW1 to AllPrep DNA spin column from step 7. Centrifuge 15 s at 11,000 rpm. Discard flow-through.

17. Add 500 µL Buffer AW2, centrifuge 2 min at max speed.

18. Place AllPrep DNA spin column in new 1.5-mL collection tube.

19. Add 40–50 µL Buffer EB directly to the AllPrep DNA spin column membrane and incubate 1 min at room temperature. Centrifuge 1 min at 11,000 rpm to elute the DNA.

20. Repeat step 19. Discard the column. The resulting 80–100 µL DNA can be used in downstream analyses (quantification, STR profiling).

3.3. RNA Quantification (see Note 9)

Quant-iT RiboGreen RNA Kit:

1. Prepare fresh RiboGreen Reagent working solution (see Note 10): dilute concentrated DMSO stock solution (Component A) 1:200 in TE Buffer.

2. Quantification mix (per sample):

1× TE Buffer	98 µL
DNase I-treated RNA extract	2 µL
RiboGreen Reagent working solution	100 µL
Total	*200 µL*

3. Mix well and incubate for 2–5 min at room temperature, protected from light.

4. Measure the sample fluorescence using a microplate reader equipped for fluorescence measurements (excitation 480 nm, emission 520 nm).

5. The RNA concentrations in the samples are calculated using an appropriate standard curve (16S and 23S ribosomal RNA) as described by the manufacturer.

Quant-iT RNA Assay:

1. Prepare fresh Quant-iT working solution (see Note 10): dilute Quant-iT RNA reagent (Component A) 1:200 in Quant-iT RNA Buffer (Component B).

2. Prepare the standards (standard 1: 0 ng/μL, standard 2: 10 ng/μL) and samples as follows:

Standards	Per sample	RNA Extract	Per sample
Quant-iT working solution	190 μL	Quant-iT working solution	198 μL
Standard	10 μL	DNase I-treated RNA extract	2 μL
Total	*200 μL*	*Total*	*200 μL*

3. Mix well and incubate for 2 min at room temperature, protected from light.

4. Run new calibration on the Qubit fluorometer using the two standards.

5. Measure the sample fluorescence.

3.4. Reverse Transcription (Superscript III First-Strand Synthesis System, Invitrogen)

Half of the RNA is used as RT minus control (without reverse transcriptase) to detect genomic DNA contamination.

1. Place 1–10 ng RNA (max. 16 μL, see Note 11) in a 0.2 or 0.5 mL tube (see Note 2).

2. Add 2 μL random hexamer (50 ng/μL) and 2 μL 10 mM dNTPs.

3. Incubate at 65°C for 5 min, place on ice for 1 min, and briefly centrifuge.

4. Prepare the following master mix (per sample): 4 μL 10× RT Buffer, 8 μL 25 mM $MgCl_2$, 4 μL 0.1 M DTT, 2 μL RNase Out (40 U/μL).

5. Add 18 μL master mix to each sample, mix gently, and briefly centrifuge.

6. RT+: pipet 19 μL into new tube and add 1 μL Superscript III RT. RT–: add 1 μL H_2O to remaining 19 μL in original tube. Mix gently and briefly centrifuge.

7. Allow to stand at room temperature for 10 min. Incubate at 50°C for 50 min, then inactivate at 85°C for 5 min, either in a thermomixer or a thermocycler. Briefly centrifuge.

3.5. PCR

Example protocol with Multiplex PCR Kit (Qiagen):

1. Primer concentrations should be optimized within the range of 0.1–0.4 μM, starting with equal concentrations (0.2 μM) of all primers.
2. PCR reaction mix (per sample):

2× Qiagen Multiplex PCR MM	12.5 μL
Primer-Mix	2.5 μL
Optional: Q-Solution (see Note 12)	2.5 μL
cDNA/H$_2$O	7.5 μL
Total	*25 μL*

3. Amplification conditions (GeneAmp PCR System 9700, AB):
   - Initial denaturation at 95°C for 15 min.
   - 30–35 Cycles of 94°C 30 s, 57°C* 90 s, 72°C 60 s (*+0.2°C/cycle).
   - Final elongation at 72°C for 60 min.

Example singleplex protocol:

1. Primer concentrations should be optimized (we typically use primer concentrations from 0.04 to 0.8 μM).
2. PCR reaction mix (per sample):

10× PCR Buffer I	2.5 μL
dNTPs (2.5 mM each)	1.25 μL
Primer-Mix	2 μL
Taq-Pol. 5 U/μL	0.25
cDNA/H$_2$O	17 μL
Total	*25 μL*

3. Amplification conditions (GeneAmp PCR System 9700, AB):
   - Initial denaturation at 95°C for 11 min.
   - 30–35 Cycles of 94°C 20 s, 55°C 30 s, 72°C 40 s.
   - Final elongation at 72°C for 60 min.

3.6. Post-PCR Purification with the MinElute PCR Purification Kit

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 flow-through.

3. Add 750 μL Buffer PE to the MinElute column and centrifuge for 1 min at 13,000 rpm. Discard 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.

3.7. Capillary Electrophoresis

1 μL of the amplified sample is added to 12.5 μL Hi-Di Formamide and 0.5 μL of GeneScan-500 LIZ or ROX Size Standard. Performance-optimized polymers POP-4, POP-6, POP-7 have been used successfully for electrophoresis (19, 23). Standard, multicolor fluorescent genetic analyzers (AB 310, 3100, 3130) and standard electrophoresis conditions (see Note 13) can be used for the detection of the body fluid-specific amplicons. Raw data are analyzed with the Genemapper Software (AB). We use an analytical threshold of 100 RFUs in determining positive results.

4. Notes

1. The use of Carrier-RNA is recommended when working with small RNA amounts.

2. DNA LoBind tubes reduce loss of genetic material caused by interaction with the plastic surface of the tube.

3. Instead of using spin baskets, one also can press/squeeze the stain with the pipette tip (caution because of foam building!), then transfer the liquid into a new LoBind tube or AllPrep DNA spin column.

4. Be sure not to take up any phenol, because this may inhibit subsequent RT or PCR reactions.

5. GlycoBlue will precipitate with the nucleic acids, facilitating good RNA or DNA recovery while increasing the size and visibility of the pellet.

6. RNAsecure Resuspension Solution is designed for direct resuspension of precipitated RNA and is capable of inactivating trace amounts of RNase.

7. DNase treatment of RNA extracts is recommended, to avoid primer and nucleotide titration by contaminating DNA. In our experience, the TURBO DNA-free Kit removes contaminating

DNA more efficiently than the on-column RNase-Free DNase Set.

8. Up to 3 h incubation at 56°C is recommended for small, old, or compromised stains.

9. Quantification with the fluorescence-based systems shows variable or negative results with small stains that were extracted with Qiagen kits. Therefore, results should be interpreted accordingly. A negative result does not indicate the inability to obtain positive downstream results for low quantity samples.

10. Use the working solution within a few hours of their preparation.

11. RNA input amount into RT should not exceed 20% of the RT volume, when DNase digestion with the TURBO DNA-free kit was applied, because components from the TURBO DNase Buffer and the DNase Inactivation reagent could interfere with the reaction.

12. Q-Solution changes the melting behavior of DNA and can improve a suboptimal PCR caused by templates that have a high degree of secondary structure or that are GC-rich.

13. Electrophoresis conditions for 3100 Genetic Analyzer: 10 s injection time, 3 kV injection voltage, 15 kV run voltage, 60°C, 21 min run time, Dye Set G5 (6-FAM, VIC, NED, PET, LIZ) or F (5-FAM, JOE, NED, ROX).

References

1. Bauer M, Kraus A, Patzelt D (1999) Detection of epithelial cells in dried blood stains by reverse transcriptase-polymerase chain reaction. J. Forensic Sci. 44:1232–1236.

2. Bauer M, Patzelt D (2002) Evaluation of mRNA markers for the identification of menstrual blood. J. Forensic Sci. 47:1278–1282.

3. Bauer M, Patzelt D (2003) Protamine mRNA as molecular marker for spermatozoa in semen stains. Int. J. Legal Med. 117:175–179.

4. Juusola J, Ballantyne J (2003) Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification. Forensic Sci. Int. 135:85–96.

5. Alvarez M, Juusola J, Ballantyne J (2004) An mRNA co-isolation method for forensic casework samples. Analytical Biochemistry 335: 289–298.

6. Ferri G, Bini C, Ceccardi S, Pelotti S (2004) Successful identification of two years old menstrual bloodstain by using MMP-11 shorter amplicons. J. Forensic Sci. 49:1387.

7. Juusola J, Ballantyne J (2005) Multiplex mRNA profiling for the identification of body fluids. Forensic Sci. Int. 152:1–12.

8. Nussbaumer C, Gharehbaghi-Schnell E, Korschineck I (2006) Messenger RNA profiling: A novel method for body fluid identification by Real-time PCR. Forensic Sci. Int. 157:181–186.

9. Fang R, Manohar CF, Shulse C, Brevnov M, Wong A, Petrauskene OV, Brzoska P, Furtado MR (2006) Real-time PCR assays for the detection of tissue and body fluid specific mRNAs. International Congress Series 1288:685–687.

10. Noreault-Conti TL, Buel E (2007) The use of real-time PCR for forensic stain identification. Promega Profiles in DNA Vol. 10/1:3–5.

11. Juusola J, Ballantyne J (2007) mRNA profiling for body fluid identification by multiplex quantitative RT-PCR. J. Forensic Sci. 52:1252–1262.

12. Bauer M (2007) RNA in forensic science. Forensic Sci. Int. Genetics 1:69–74.

13. Bauer M, Patzelt D (2008) Identification of menstrual blood by real time RT-PCR: Technical improvements and the practical value of negative results. Forensic Sci. Int. 174:54–58.

14. Zubakov D, Hanekamp E, Kokshoorn M, van Ijcken W, Kayser M (2008) Stable RNA markers for identification of blood and saliva stains from whole genome expression analysis

of time-wise degraded samples. Int. J. Legal Med. 122:135–142.

15. 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.

16. 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. Genetics 3:80–88.

17. Haas C, Muheim C, Kratzer A, Bär W, Maake C (2009) mRNA profiling for the identification of sperm and seminal plasma. Forensic Sci. Int. Genetics Supplement Series 2:534–535.

18. Nakanishi H, Kido A, Ohmori T, Takada A, Hara M, Adachi N, Saito K (2009) A novel method for the identification of saliva by detecting oral streptococci using PCR. Forensic Sci. Int. 183:20–23.

19. Haas C, Hanson E, Bär W, Banemann R, Bento AM, Berti A, Borges E, Bouakaze C, Carracedo A, Carvalho M, Choma A, Dötsch M, Durianciková M, Hoff-Olsen P, Hohoff C, Johansen P, Lindenbergh PA, Loddenkötter B, Ludes B, Maroñas O, Morling N, Niederstätter H, Parson W, Patel G, Popielarz C, Salata E, Schneider PM, Sijen T, Sviezená B, Zatkalíková L, Ballantyne J (2011) mRNA profiling for the identification of blood—Results of a colla-borative EDNAP exercise. Forensic Sci. Int. Genetics. 5(1):21–26.

20. Fleming RI, Harbison S (2010) The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids. Forensic Sci. Int. Genetics 4:244–256.

21. Fleming RI, Harbison S (2010) The use of bacteria for the identification of vaginal secretions. Forensic Sci. Int. Genetics 4:311–315.

22. Haas C, Hanson E, Kratzer A, Bär W, Ballantyne J (2011) Selection of highly specific and sensitive mRNA biomarkers for the identification of blood. Forensic Sci. Int. Genetics. 5(5): 449–458.

23. Haas C, Hanson E, Anjos MJ, Bär W, Banemann R, Berti A, Borges E, Bouakaze C, Carracedo A, Carvalho M, Castella V, Choma A, De Cock G, Dötsch M, Hoff-Olsen P, Johansen P, Kohlmeier F, Lindenbergh PA, Ludes B, Maroñas O, Moore D, Morerod ML, Morling N, Niederstätter H, Noel F, Parson W, Patel G, Popielarz C, Salata E, Schneider PM, Sijen T, Sviežena B, Turanská M, Zatkalíková L, Ballantyne J (2011) RNA/DNA co-analysis from blood stains – results of a second collaborative EDNAP exercise. Forensic Sci. Int. Genetics. doi: 10.1016/j.fsigen.2011.02.004.

24. Lee J, Hever A, Willhite D, Zlotnik A, Hevezi P (2005) Effects of RNA degradation on gene expression analysis of human post mortem tissues. FASEB J. 19/10:1356– 1358.

25. Bauer M, Polzin S, Patzelt D (2003) Quantification of RNA degradation by semi-quantitative duplex and competitive RT-PCR: a possible indicator of the age of bloodstains? Forensic Sci. Int. 138:94–103.

26. 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.

27. Zubakov D, Kokshoorn M, Kloosterman A, Kayser M (2009) New markers for old stains: stable mRNA markers for blood and saliva identification from up to 16-year-old stains. Int. J. Legal Med. 123:71–74.

28. Smith PJ, Ballantyne J (2007) Simplified low-copy-number DNA analysis by post-PCR purification. J. Forensic Sci. 52:820–829.

29. Juusola J (2005) Messenger RNA profiling: A prototype method for body fluid and tissue identification. Doctoral thesis, University of Central Florida.

Chapter 13

DNA Extraction from Aged Skeletal Samples for STR Typing by Capillary Electrophoresis

René Huel, Sylvain Amory, Ana Bilić, Stojko Vidović, Edin Jasaragić, and Thomas J. Parsons

Abstract

STR analysis of DNA extracted from skeletal samples can play an important role in the identification of missing persons. Here we present a method for the extraction of DNA from skeletal samples involving complete demineralization and digestion of the sample, followed by purification by silica binding. This method, together with the multiplex STR typing approach also presented, has proven highly successful in the recovery of DNA profiles from degraded, aged skeletal remains from a wide range of environmental contexts. The methodological steps presented include bone decontamination and grinding, DNA extraction, repurification in the case of highly inhibited samples, quantification, STR multiplex amplification, and profile reporting guidelines. However, the conditions applied for amplification and the criteria for allele calling and profile submission must be based on the results of each laboratory's internal validation experiments involving the type of samples relevant to the project at hand. The methods presented here have permitted large-scale DNA-based identification of persons missing from mass disasters and armed conflict.

Key words: DNA extraction, Bone, Demineralization, Silica, STR Analysis, Missing persons' identification

1. Introduction

The use of nuclear STR typing for human identification has become well established as a primary tool in victim identification from a wide range of contexts (1), including natural disasters (e.g., the 2004 SE Asian tsunami, and Hurricane Katrina), terrorist attacks (e.g., the September 11 attacks), numerous transportation accidents, and armed conflict/human rights violations. DNA typing is unique among identification methods in having a rigorously established

scientific basis (2) and often permits a reliable means for identification when other methods cannot be applied. This is often the case in instances of body fragmentation where the postmortem samples do not retain identifiable characteristics, or when medical, dental, or fingerprint data are not available for the missing persons.

The protocols presented here are based on current, validated standard operating procedures of the International Commission on Missing Persons (ICMP), developed in the course of the world's largest missing persons' identification project. The ICMP was established to assist in addressing the issue of ~40,000 persons missing as the result of armed conflict and human rights violations in the former Yugoslavia in the 1990s ((3); www.ic-mp.org). Since 2001, the ICMP has generated nuclear autosomal STR profiles from over 33,000 bone and tooth samples, primarily from mass graves in the former Yugoslavia, or from victims of natural disasters (2004 SE Asian tsunami, 2005 hurricane Katrina, and 2008 Typhoon Frank in the Philippines). Genetic kinship comparisons between victim DNA profiles and those of surviving relatives of victims have produced over 29,000 highly statistically significant DNA matches, representing over 16,500 missing individuals and numerous additional reassociations of fragmented remains.

Innumerable methods have been published for DNA extraction from bone, many coming from the ancient DNA research community (e.g., see (4)) and others established in applied forensic laboratories. For many years, aged or degraded skeletal remains (e.g., buried for years in the soil) were considered mainly subjects for mtDNA testing, due to the higher chance for success with this multicopy target. In forensics, the most commonly applied extraction methods for bone involved proteinase K digestion followed by phenol/chloroform extraction (e.g., (5)). Early trials in forensic protocols showed no advantage to demineralization of bone samples by extensive EDTA washes (6), a commonly used step in ancient DNA protocols. In setting out to apply nuclear STR typing to aged bone samples on a large scale, the ICMP initially did not have good success with standard phenol/chloroform methods and developed a silica based method for large volume bone extractions, which gave much better results and permitted thousands of nuclear STR-based identifications to be made (7, 8).

The survival of DNA in bone is based on its protective mineral and protein matrix, which serves as mechanical barrier to microorganismal attack and at least partially isolates the DNA from detrimental conditions of the external environment. DNA typing success rates are best with dense cortical bone from weight-bearing elements (9–11). DNA appears to be best protected in small, dense crystalline aggregates within the bone (12), suggesting the concept that DNA extraction methods should be optimized to ensure access to densely mineralized "preferred niches" within the sample. To this end, the benefit of demineralization steps was revisited in a

systematic manner (13) demonstrating a greater of DNA when demineralization is accomplished without discarding the EDTA solution used to effect decalcification. The DNA extraction method presented here combines the benefits of complete demineralization with a convenient silica purification method (14) that is comparatively effective at removing PCR inhibitors. While using smaller amounts of sample, this method provides greater success rates with difficult samples than ICMP's previous method. This method is similar to that of other recent publications (15, 16). The method has been validated in accordance with ISO 17025 accreditation standards and has successfully been applied to thousands of missing persons' casework samples.

2. Materials

Separate laboratory facilities are required for pre-PCR (sample preparation, bone sample washing and grinding, DNA extraction, quantification, and PCR amplification setup) and post-PCR (product purification and capillary electrophoresis) to avoid PCR product contamination. Bone sanding and grinding require ventilated hoods or other dust collection apparatus to capture/remove the bone powder that is released into the air during the process; these steps are best performed in a separate laboratory from other pre-PCR steps. Ear protection is recommended for bone grinding. For personal protection and contamination avoidance, the following should be used throughout: clean or disposable lab coats and pants, hair nets, face masks, gloves, disposable sleeves to cover wrists, and disposable overshoe covers. Proper hoods should be used for DNA extraction and PCR setup (the ICMP uses dead-space hoods, but other laboratories employ biofilter hoods for similar work).

2.1. Sanding the Sample

1. Dremmel or dental type rotary sanding tool with a variety of sanding bits to accommodate different sizes and shapes of bone samples, with foot pedal operated variable power control.

2.2. Washing/Grinding the Sample

1. Tweezers/forceps
2. Waring blender cups, model SS110
3. Waring blender base, 1 L, model 7009 or 7011
4. 50 mL conical tubes
5. Commercial Bleach (5% sodium hypochlorite) (see Note 1)
6. Ethanol 96%

2.3. Extraction/ Purification

1. Demineralization buffer: Ultrapure 0.5 M EDTA pH 8.0 (Gibco, Invitrogen Corporation, Carlsbad, CA), 1% Sodium-*N*-laurylsarcosinate (Fluka BioChemika, Buchs, Switzerland).

Add 1 g of Sodium-*N*-laurylsarcosinate per 100 mL of EDTA. Mix gently until the Sodium-*N*-laurylsarcosinate is completely dissolved. Store unused portions at 4°C for up to 1 month (see Note 2).

2. Amicon Ultra-15–100,000NMWL (Millipore Corporation, Bedford, MA) (see Note 3).

3. Qiagen QIAquick PCR purification kit (Qiagen Inc., Valencia, CA) contains:

    (a) Qiaquick spin columns.

    (b) Buffer PB.

    (c) Buffer PE (note: ensure ethanol is added according to the manufacturer's instructions).

    (d) Buffer EB (see Note 4).

4. Proteinase K, 10 mg/mL (Promega Corporation, Madison, WI).

5. Ethanol 96%.

3. Methods

Ethanol is added to buffer PE prior to use. Follow directions for volumes according to the manufacturer.

Buffer PB contains chaotropic salts. Appropriate safety measures should be taken when handling this reagent.

In order to prevent contamination, all steps in this procedure should be performed in dedicated preamplification and postamplification labs as appropriate. Scientists conducting the manipulations must have an excellent awareness of the potential for contamination of exogenous DNA which can be introduced into the process at many stages by clumsy technique, such as passing sleeves over open tubes or touching the inside of tube caps when opening or closing tubes. It is not possible to detail all aspects of preventative manipulations, but it is recommended that scientists be familiar with literature pertaining to "ancient DNA" precautionary guidelines and adapt their manual technique to minimize the potential for accidental introduction of contaminating cellular material or DNA.

Any specimens suspected of containing large quantities of DNA (e.g., fresh bone) should be extracted separately from older samples.

A surgical facemask, gloves, hair cap, disposable sleeves, dedicated lab pants, and shoes (or shoe covers) must be worn at all times to prevent contamination. Personal protective clothing (PPC) should not leave the areas in which they are used.

All trays, pipettes, benches, and hoods must be thoroughly cleaned with bleach and/or UV irradiated (in a crosslinker for 30 min at maximum emission at 254 nm).

Any method, protocol, or instrument should be internally validated by the lab according to accepted forensic standards.

3.1. Sanding and Preparation of the Bone Sample

1. Sizeable bones or portions thereof should first be cut to a manageable size for subsequent processing (see Note 5). This can be done using specialized bone saws, or manually by hack saws, etc. Samples between 3 and 10 g are optimal to begin with, generally providing enough material for several extractions.

2. Sanding applies only to bone samples. Teeth do not undergo sanding. If processing tooth samples, go directly to washing and grinding.

3. Prepare each sample separately (see Note 6).

4. Before sanding, remove any adhering tissue from the bone using a scalpel and tweezers/forceps.

5. Place the sample in a holding vice to avoid the sample moving while sanding.

6. Sand the exterior (see Note 7), ensuring that the entire superficial surface of the sample is removed (1–3 mm, depending on the type of sample). Attempt to sand away any particularly discolored or porous portions of the sample, if sufficient sample exists, as dense cortical bone has the best DNA preservation in aged samples.

7. Depending on the weight of the sample, cut the sample in half to retain reserve material if needed. Typically, 2–3 g is enough for several extractions.

8. Thoroughly clean all surfaces, equipment, and the interior of the hood with bleach between each sample.

3.2. Washing and Grinding the Bone Sample

1. If the bone/tooth sample becomes moldy as a result of storage after sanding, it should be washed in tap water and the mold mechanically removed with forceps or rubbing, prior to the washing procedure. Teeth should be thoroughly cleaned under running water and any dirt or adhering material removed using forceps.

2. Unless the sample is less than 1 g, it will likely be desirable to break the sample into smaller pieces of ~0.5–1 g to facilitate subsequent steps of washing and grinding (see Note 8). If the sample is a tooth, move directly to step 4.

3. Weigh the sample. Typically, only 2–3 g is necessary for several extractions. Place 2–3 g (or as much as required) into a new 50 mL tube. The remaining sample can be kept in reserve if needed in the future.

4. Wash samples as follows (see Note 9):

- Each wash is discarded before the next wash step is undertaken.

- Two washes with a 10% dilution of commercial bleach (see Note 1), immersing the sample completely with vigorous shaking for 30 s.

- Two washes in deioinized water, where available, or tap water, with the sample fully immersed with vigorous shaking for 30 s. Following vigorous shaking for the second wash, the sample is to be left submerged in water for a minimum of 5 min before the wash discarded.

- Two washes in 96% ethanol with vigorous shaking for 20–30 s.

5. Place washed samples in open tubes in an oven/incubator set between 50 and 60°C and allow samples to completely dry (see Note 10).

6. Remove samples from oven/incubator when drying is complete.

7. Place the sample fragments into the blender cup, cover it with the lid and secure it on the blender base (see Note 11). Run the blender for no more than 2 min, then check if the sample is finely ground. The sample can be transferred to a sterile weighing dish for closer inspection. Larger fragments are to be returned to the blender cup for further grinding until a fine powder is obtained (see Note 12).

8. Transfer the final powder from the blender cup into a weighing dish, then transfer the powder into a new (prelabeled) 50-mL tube.

9. Clean blender cups, base, and hood well between each sample (see Note 13).

3.3. Extraction and Purification of DNA from the Bone Sample

One reagent blank (extraction negative control) should be carried throughout the extraction procedure and assayed in parallel with the test samples (see Note 14).

1. Preheat oven/incubator to 56°C.

2. Weigh between 0.5 and 1.0 g of bone powder into a new 50-mL tube (see Note 15).

3. To each extraction tube, carefully add 15 mL of digestion buffer (see NOTES 16 and 17).

4. Repeat for each sample extraction tube, with the final tube being the extraction negative control.

5. Add 1 mL of Proteinase K (10 mg/mL) to each extraction tube, including the extraction negative control.

6. Ensure the lid of each tube is tightly secured. Wrap each one with Parafilm to ensure no leaks occur during incubation.

7. Place samples horizontally onto the rocking platform located in the incubator/oven (see Note 18). Set the rocking speed to ~90 rpm (enough to ensure that the liquid and powder are in contact with each other and that they are mixing).

8. Incubate samples overnight (typically all bone material is digested within 5–6 h).

9. The following day, remove the tubes from the incubator. Ensure that all bone material is completely digested.

10. If significant undigested material is remaining, check the level of demineralization buffer to ensure no leaking occurred during the incubation. If leaking occurred, add additional demineralization buffer up to 15 mL and add another 1 mL of Proteinase K (10 mg/mL). Place the sample(s) back on the rocking platform and continue incubation for another 5 h (see Note 19).

11. Place tubes into a centrifuge with a swinging bucket rotor and 50 mL adaptors, and spin samples for 5 min at 3,000 rpm to pellet any small amounts of remaining bone debris.

12. Once the centrifugation is complete, carefully transfer the supernatant from each sample tube into a corresponding prelabeled 15-mL centrifugal filter unit.

13. Place the Amicon Ultra-15 units back into the centrifuge and spin for 30 min at 4,000 rpm.

14. Prelabel corresponding 1.5 and 2 mL tubes for each sample plus the negative control. Preload the 2 mL tubes with 1.5 mL of PB buffer.

15. After centrifugation, place the Amicon Ultra-15 units back inside the hood. Verify that the volume of lysate remaining on the filter of the Amicon Ultra-15 is between 300 and 450 μL:

    (a) If it is the case, use a 200 μL pipette to transfer the volume into a 2-mL tube containing PB buffer.

    (b) In case the Amicon Ultra-15 contains less than 300 μL of lysate, add water up to 300 μL before continuing.

    (c) In case the Amicon Ultra-15 containing the extraction negative has been dried out, add 300 μL of water and transfer it to the PB tube after 3–5 min.

    (d) If the Amicon Ultra-15 contains more than 450 μL, replace the tube in the centrifuge and spin it at 4,000 rpm until a volume less than 450 μl has been reached. Transfer to a 2-mL tube preloaded with 1.5-mL of PB buffer.

16. Mix each sample and buffer PB well by inverting the tubes several times. Spin the tubes very briefly in a centrifuge prior to the next step.

17. Transfer approximately 630 µL of PB/sample to a prelabeled corresponding QIAquick column and spin for 1 min at 13,000 rpm in a fixed rotor centrifuge.

18. Discard the flow-through and repeat step 17 until the volume of each sample has transferred.

19. Wash each column with 750 µL of buffer PE (ensure the recommended amount of ethanol has been added to the buffer) and spin again for 1 min at 13,000 rpm.

20. Discard the flow-through and repeat the wash step 2 additional times, each time discarding the flow-through, for a total of three washes.

21. Remove any residual ethanol from the final wash by spinning each column for 3 min at 13,000 rpm.

22. Place the QIAquick column on a new, and labeled, 1.5 mL tube. Add 50 µL of buffer EB to each column membrane and allow the buffer to incubate on the membrane for 1 min.

23. Elute the DNA from the column by centrifuging samples for 1 min at 13,000 rpm.

24. Store samples at –20°C.

3.4. Repurification of Inhibited DNA Extracts

Some DNA extracts may have inhibitors copurify during extraction which can affect downstream STR analysis. These can be detected in some quantification methods (delayed Ct IPC) or in the STR profile itself. In extreme cases, this can be seen as coloration in the DNA extract itself. To reduce the level of inhibitors in such extracts, use the simple protocol below to repurify these samples (see Note 20).

One reagent blank should be carried throughout the repurification procedure and assayed in parallel with the samples.

1. To each sample selected for repurification, add 5 volumes of buffer PB to the volume of DNA extract and mix the PB buffer and the DNA extract by turning the tube up and down ~10 times, ensuring the contents are mixed well.

2. From each 1.5 mL tube, transfer all the PB buffer/DNA extract into the QIAquick spin column. Place the tubes into a fixed angle centrifuge and spin for 1 min at 13,000 rpm.

3. After centrifugation, place the tubes back on the rack and discard the flow-through.

4. Add 750 µL of PE buffer (ensure EtOH has been previously added to the buffer) to each QIAquick spin column. Place the tubes into the centrifuge and spin for 1 min at 13,000 rpm.

5. After centrifugation, place the QIAquick spin columns back on the rack and discard the flow-through.

6. Repeat steps 4–5 two additional times for a total of three washes.

7. To dry the column completely, place the tubes in the centrifuge and spin for 3 min at 13,000 rpm.

8. Place the columns back on the rack and inside the hood transfer the filter of the QIAquick spin column in a clean, labeled 1.5 mL tube.

9. Add 50 µL of EB buffer (see Note 21) directly on the center of the QIAquick membrane and let it stand for 1 min. Place the elution tubes + QIAquick column filters into the centrifuge and spin for 1 min at 13,000rpm.

10. Place the elution tubes + QIAquick column filters back on the rack, verify that the elution tubes contain approximately 50 µL of DNA extract (or whatever volume is being eluted) and throw away the QIAquick column filters and close the elution tube.

11. Store the tubes at –20°C.

3.5. Quantification, Amplification, Fragment separation, and Analysis

1. DNA extracts from bone or tooth samples should be quantified prior to STR analysis due to the varying amount of DNA that can be obtained due to differences in the type of skeletal element, the age and environmental origin of the samples, and the possibility of coextraction of inhibitors. The ICMP uses Applied Biosystems' Quantifiler Human DNA quantification kit on all bone samples. Samples exhibiting a Ct IPC greater than 30 are repurified using the protocol listed above. These DNA extracts are then requantified to determine if the inhibition is still present. It is possible to repurify a DNA extract using this method up to 2 times.

2. For DNA extracts with a yield >1ng/µL, an aliquot is diluted to 1 ng/µL and 1 µL is used in PCR. For all samples with a DNA yield between 0.5 and 1 ng/µL, aliquots of 1 and 2 µL are used in subsequent PCR amplifications using Promega's PowerPlex16. For samples with <0.5 ng/µL, aliquots of 2 µL and 5 µL are used for PCR (see Note 22).

3. Amplify samples using Promega's PowerPlex16 (see Note 23). Mastermix is set up according to the manufacturer's recommended volumes except an additional 0.2 µL of *taq* is added per reaction. Cycling parameters are listed in Fig. 1 and are carried out on a Applied Biosystems 9700 GeneAmp thermal cycler.

4. Capillary electrophoresis of samples should follow manufacturer's recommendations, with possible additional modifications based on results of internal validation studies of particular instruments. For Promega's PowerPlex16 STR kit, the ICMP

Temperature	Time	Ramp Rate	
95°C	11 minutes	100%	
96°C	1 minute	100%	
94°C	30 seconds	100%	10 cycles
60°C	30 seconds	29%	
70°C	1.5 minutes	23%	
90°C	30 seconds	100%	24 cycles
60°C	30 seconds	29%	
70°C	1.5 minutes	23%	
60°C	45 minutes	100%	
4°C	hold indefinitely		

Fig. 1. Modified cycling parameters for PowerPlex16 amplification of bone samples on a ABI 9700 GeneAmp thermal cycler.

uses the manufacturer's recommended volumes of reagents with a 3 kV injection for 11 s on an Applied Biosystems' 3130xl Genetic Analyzer (see Note 24).

5. A minimum of two amplifications are performed on each sample. As quantification kits may not always give an accurate assessment of DNA quantity or quality, carefully review initial amplifications and alter amplification strategies according to the resulting profile(s). For example, if inhibition is suspected at a certain volume of template, try amplifying with half the original input.

6. Reporting of alleles in STR typing should be based on internal validation experiments (addressing stutter, peak balance, minimum allele calling thresholds, etc.). Based on ICMP's validation, the following general guidelines are used: Profiles are considered reportable if results are obtained from a minimum of 11 STR loci and Amelogenin; homozygotes are called with a minimum RFU threshold of 300, and heterozygotes are called with a minimum rfu threshold of 50; alleles are reported only if they are reproduced in a minimum of two different amplifications. In addition to these guidelines, an overall conservative approach is taken, and in many instances, duplicate extractions are performed to confirm results or to obtain additional reportable alleles (see Note 25).

4. Notes

1. For work space and equipment decontamination, and for bone washing, a 10% solution of commercial bleach is used (0.5% sodium hypoclorite). Fresh solutions of 10% bleach should be made daily.

2. Purchasing ultrapure 0.5 M EDTA, pH 8.0 saves effort in the preparation and the pH balancing of this solution.

3. Different molecular weight cutoffs are available. Some laboratories use smaller size cutoffs (30,000 or 50,000); however, we have not seen a difference in results between 100,000 and 50,000 except for longer centrifugation times.

4. The reagents included in one kit (250 reactions) are enough for ~100 extractions.

5. The ICMP has detailed protocols for selecting the optimum sample from various skeletal remains assemblages and can be obtained from the authors upon request.

6. It is good practice to photodocument the sample before it undergoes processing. This way there is a visual historical record of the sample. If the original sample code is changed to conform to a laboratory numbering system, it is very beneficial to photodocument the sample with its original code and its new laboratory code.

7. A variety of sanding attachments can be used. Having various sizes and shapes will ensure that all types of samples can be efficiently sanded.

8. Any alternate means of subdividing the bone sample is acceptable, such as with a dremel cutting bit, etc.

9. For fresh samples (<1-year postmortem), double the amount of washes in each step to reduce the amount of potential external contamination. In such cases, four washes with 10% bleach, four water washes, and four ethanol washes are recommended.

10. Drying times vary depending on the sample. Fresh samples may require additional time to completely dry.

11. Other ways of grinding samples may be used, including liquid nitrogen freezer mills and powdering a sample using a drill. With complete demineralization, there is no advantage to extremely finely ground powder.

12. If many samples are being processed, at once have additional cups and lids available.

13. Blender cups and lids are to be cleaned as follows: After grinding, add a small volume of 1% Liquinox to the blender cup and seal with lid. Place on the blender base and run for 10–20 s to rinse the cup and lid. Discard the Liquinox, then place the blender cups in the sink to soak in 1% Liquinox, until ready to proceed to the next step. Thoroughly clean the blender cups and lids using a sponge and hot water making sure to wash the area under the blade. Rinse the cups and lids with tap water. Soak them in 10% bleach (in a plastic container) for at least 30 min. Rinse with deionized water (or tap water if deionized is unavailable) and then 96% ethanol and leave them to air dry.

If another method is used to grind samples, ensure that the equipment is thoroughly decontaminated between samples.

14. Throughout the procedure, sample transfers between different tubes should be witness checked by an independent person. This will reduce the likelihood of sample switches.

15. Less than 0.5 g can be used in case of small samples. No more than 1 g of bone powder should be added to 15 mL of digestion buffer since 15 mL of 0.5 M EDTA can theoretically bind only the amount of calcium contained in ~1 g of bone.

16. For large numbers of extractions, prepare the demineralization buffer in a large glass bottle and use a large dispensette, set to 15 mL, to accurately pipette the digestion buffer into tubes.

17. Typically 11 samples and 1 extraction negative control can be extracted by one person at one time. This keeps samples balanced in the centrifuge while also being a manageable amount of samples to extract per analyst.

18. Various ways of shaking/incubation can be used. Constant rocking motion where the bone powder and digestion buffer are constantly mixing is optimal.

19. If substantial undigested material still remains and no leaking occurred, pellet the undigested material as in step 11 and retain the pellet. Add another 15 mL of demineralization buffer to the tube along with another 1 mL of Proteinase K (10 mg/mL) and incubate again overnight. The resulting fractions can be processed individually or can be pooled together in step 12.

20. The repurification protocol also works well on samples that have been extracted using other protocols.

21. For samples with low amounts of DNA, one can elute in a 30 μL volume.

22. Commercial kits such as Applied Biosystems Quantifiler Human DNA kit, or similar type, may not detect all inhibitors present in DNA extracts from bones. For this reason, 2 and 5 μL are used as an initial amplification of most samples. Based on the resulting profiles, additional amplifications can be adjusted accordingly. It is sometimes the case that using less DNA extract provides better results even when quantification shows little DNA and no evidence of strong inhibition.

23. Any validated multiplex STR kit may be used with this extraction method. A wide range of commercial kits are available with different buffer formulations and STR amplicon sizes that can affect typing success depending on the quality and quantity of DNA obtained. The number of loci that should be targeted depends on the nature of the case/event. Events involving a small number of people, and/or with reference samples directly from the missing persons themselves, will not

require a large number of loci for highly significant results. Larger events with many missing persons, particularly if reference samples from relatives are used, call for a greater number of loci. Depending on the reference samples that are available, multiple different multiplex kits may be needed to provide enough loci for sufficiently significant results from genetic kinship analysis.

24. Sensitivities between instruments may vary within a lab, and from lab to lab. Manufacturers may recommend a range for both injection voltage and injection time that should be tested in internal validation experiments.

25. DNA extractions from aged bone samples are subject to a wide range of copurifying molecules. In addition to a potentially wide range of PCR inhibitors with a variety of effects, often substantial quantities of microbial or fungal DNA are also extracted. The latter increase the sequence complexity of the PCR reaction and can decrease PCR efficiency. In some cases, nonspecific amplification products of microbial origin can be obtained. These peaks usually have different morphology and do not correspond to size bins of authentic amplicons, but it is important that analysts be aware of the potential for such artifacts.

Acknowledgments

References

1. Alonso A, Martin P, Albarran C et al (2005) Challenges of DNA profiling in mass disaster investigations. Croat Med J 46:540–548.
2. National Research Council of the National Academies (2009) Strengthening forensic science in the United States: a path forward. National Academies Press, Washington DC.
3. Huffine E, Crews J, Kennedy B et al (2001) Mass identification of persons missing from the break-up of the former Yugoslavia: structure, function, and role of the International Commission on Missing Persons. Croat Med J 42:271–275.
4. Rohland N and Hofreiter M (2007) Comparison and optimization of ancient DNA extraction. Biotechniques 42:343–352.
5. Holland MM, Fisher DL, Mitchell LG (1993) Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. J Forensic Sci 38:542–553.
6. Fisher DL, Holland MM, Mitchell L (1993) Extraction, evaluation, and amplification of DNA from decalcified and undecalcified United States Civil War bone. J Forensic Sci 38:60–68.

7. Davoren J, Vanek D, Konjhodzić R et al (2007) Highly effective DNA extraction method for nuclear short tandem repeat testing of skeletal remains from mass graves. Croat Med J 48:478–485.

8. Parsons TJ, Huel R, Davoren J et al (2007) Application of novel 'mini-amplicon' STR multiplexes to high volume casework on degraded skeletal remains. FSI:Genetics 1:175–179.

9. Miloš A, Selmanović A, Smajlović et al (2007) Success rates of nuclear short tandem repeat typing from different skeletal elements. Croat Med J 48:486–493.

10. Mundorff AZ, Bartelink EJ and 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.

11. Edson SM, Ross JP, Michael D. Coble MD et al (2004) Naming the dead: confronting the realities of the rapid identification of degraded skeletal remains. Forensic Sci Rev 16:63–90.

12. Salamon M, Tuross N, Arensburg B et al (2005) Relatively well preserved DNA is present in the crystal aggregates of fossil bones. Proc Natl Acad Sci 102:13783–13788.

13. Loreille OM, Diegoli TM, Irwin JA et al (2007) High efficiency extraction of DNA from bone by total demineralization. Forensic Sci Int Genet 1: 191–195.

14. Yang DY, Eng B, Waye JS et al (1998) Technical note: improved DNA extraction from ancient bones using silica-based spin columns. Am J Phys Anthropol 105:539–543.

15. Lee HY, Park MJ, Kim NY et al (2010) Simple and highly effective DNA extraction methods from old skeletal remains using silica columns. Forensic Sci Int Genet 4: 275–280.

16. Seo SB, Zhang A, Kim HY (2010) Technical note: efficiency of total demineralization and ion-exchange column for DNA extraction from bone. Am J Phys Anthropol 141: 158–162.

Chapter 14

Interpretation Guidelines for Multilocus STR Forensic Profiles from Low Template DNA Samples

Zoran M. Budimlija and Theresa A. Caragine

Abstract

Low template (LT) DNA testing is a more sensitive method of PCR DNA typing which tests lower quantities of DNA compared to traditional PCR DNA protocols. Methods applied in this testing involve amplification or postamplification efforts to increase detection sensitivity. Establishing the interpretation rules of the results obtained is *condition sine qua non* for successful incorporation of this valuable technique into forensic casework. Here we describe a successfully optimized and validated approach to interpretation of LT-DNA samples.

Key words: DNA, Interpretation, Low template DNA, Optimization, Forensics

1. Introduction

Low template (LT) DNA (also known as high sensitivity DNA, low level DNA, trace DNA, touched DNA) testing is a state-of-the-art technology used to recover and detect small amounts of DNA. It has been used throughout the scientific world for many years and it is not unique specifically to the forensic community (1–3). In 1996, Taberlet et al. (4) described "reliable genotyping of samples with very low DNA quantities using PCR," while in 1997 Findlay et al. reported single cell STR analysis (5) for applications other than forensics. As early as 1999, the Forensic Science Service in the UK implemented LCN DNA in their casework as an alternative to mtDNA (6).

Low template DNA (LT-DNA) analysis can be defined as DNA typing results from samples with low input amounts and/or compromised quality which show enhanced stochastic effects. These random phenomena include allelic drop-in, allelic drop-out, increased stutter formation, and intralocus peak height imbalance.

The recognition and interpretation of these effects are based upon a laboratory's internal validation (6–10).

Template amounts that generate these increased stochastic effects typically are below 100 pg of DNA, the equivalent of 16 diploid cells (one diploid cell contains ~6 pg of DNA). LT-DNA testing is a modified version of routine (high template) nuclear DNA testing (HT-DNA). It examines the same regions of DNA and uses the same procedures, but they are adjusted to increase testing sensitivity by one or more of the following means: additional amplification cycles, postamplification purification, reduced reaction volume, capillary electrophoresis injection enhancement by increased voltage or time, and/or nested PCR. Preamplification strategies, such as reextraction or DNA template concentration are not considered LT-DNA methods (11–19).

The potential for typing inaccuracies and irreproducibility due to stochastic effects and contamination risks are notable concerns that any laboratory considering LT-DNA analysis should contemplate and evaluate (7, 20). That is why special attention must be paid to the optimization and incorporation of existing HT-DNA methodologies in the development of LT-DNA protocols. Stringent quality control and detailed testing procedures, as well as specific interpretation protocols designed to accommodate stochastic effects are necessary.

2. Materials (Quality Practices and Testing Procedures)

2.1. Quality Practices

To validate and implement LT-DNA testing, the laboratory must start with excellent quality practices, pertaining to the following:

1. Facility conditions—Space should be dedicated to LT-DNA testing, consumables and buffers should be free of DNA, analysts should don proper personal protective attire, and laboratory space and utensils should be cleaned before and after LT-DNA work (21, 22) (see Note 1).

2. Quality control tests on all reagents—All reagents used should be tested with their respective assays to ensure they are functioning and are free of DNA (see Note 2).

3. Quality control tests on all instruments (see Note 3).

4. Evaluation of controls with every test—Each test should include the appropriate controls. Positive controls test the robustness of the assay and should be amplified using an appropriate DNA input amount. Negative controls, including extraction negatives, purification negatives, and amplification negatives identify possible contamination during testing and are treated using the most sensitive conditions applied to the samples. Since it is recommended that all controls be amplified in triplicate (7), with

an extraction batch size of nine samples, for example, there may be up to nine negative control amplifications. Due to the sensitivity of the system, some alleles may be detected in the negative controls; therefore, interpretation guidelines should be established based on a laboratory's validation (see Note 4).

5. Measures to detect exogenous DNA in a sample—Efforts should be made to build an elimination database containing the DNA profiles of all personnel who could potentially have contact with the evidence or the testing facility. These profiles are for quality control purposes only.

6. Analysts qualifying tests—Samples with known DNA profiles which are unknown to the analyst should be used to perform qualifying tests. In order to mimic appropriate casework conditions, sample extracts should be diluted to LT-DNA levels, if necessary. These tests are performed following training in each procedure and at consistent intervals thereafter.

2.2. Testing Procedures

1. Evidence triage and extraction—Virtually any type of biological sample could be a candidate for the LT-DNA testing approach. However, some of the most probable substrates for this kind of testing include the following touched items (where skin cells were potentially deposited): clothing that has been touched or grabbed, handled items, car swabs, handles of brushes or combs, tools jewelry, keys, lighters/matches, pens, sides of bottle and cans, weapons (gun swabs, knife, handles, sticks), limited contact samples (door knob, window sill), fingerprints (fresh and archived), air bags, letters and envelopes, and objects used for binding or strangulation (cords, ropes, strings, zip ties, tape). Degraded and/or inhibited samples may also be candidates for this testing. Potential LT-DNA samples should be extracted using a specific protocol designed to enhance DNA recovery (see Note 5).

2. Quantitation—The determining factor as to whether the sample will be treated as a LT-DNA sample is the recovered quantity of DNA. Samples wherein less than 100 pg of template DNA would be amplified in each reaction tube are generally considered LT-DNA samples (see Note 6).

3. Amplification strategy and adjustments.
    The amplification strategy presented in this chapter is based on internal validation (7). The AmpFlSTR® Identifiler PCR Amplification Kit (Applied Biosystems) is used according to the manufacturer's recommendations with the exception of a 2-min annealing time for each of 31 cycles and the use of a half-reaction volume with 2.5 U of Taq. All amplifications are carried out in the ABI GeneAmp® PCR System 9,700 thermal cycler (Applied Biosystems) within thin-walled 0.2 mL AB MicroAmp® (Applied Biosystems) (see Note 7).

4. Separation and Analysis.

Five microliters of each PCR product and 0.375 µL of Gene-Scan® 500 LIZ® Size Standard (Applied Biosystems) is prepared with HIDI formamide (Applied Biosystems) for a total volume of 50 µL and injected with parameters established with validation studies. Since often there are large peak height differences between short and large loci, in order to detect a profile in its entirety, samples may be injected with more than one injection condition. In these cases, results for each locus from the injection of highest quality should be used (see Note 8).

In addition to injecting each sample replicate, 5 µL of each of the three replicate PCR products are combined and mixed to form a pooled sample. Despite the fact that this sample is not to be evaluated autonomously, it is very useful for consecutive analysis, due to the fact that allelic peak imbalance is less pronounced.

3. Methods (Interpretation Procedures and Guidelines)

3.1. Interpretation Procedures

3.1.1. General Considerations

Testing and interpretation protocols established through validation (7, 23), in addition to stringent quality control measures, address the concerns regarding stochastic effects observed with LT-DNA samples. For example, stutter peaks are more common, and if they occur in an early cycle of PCR, they will be larger relative to the height of true peaks. Nonrepeating alleles, otherwise termed allelic drop-ins are a very minor component of the template. It cannot be determined whether drop-ins were deposited prior to, during, or after evidence collection. Peak imbalance may arise through preferential sampling of very few copies of an allele. In extreme cases, an allele may not be amplified (drop-out) resulting in false homozygosity or, even locus drop-out.

3.1.2. Consensus Approach

The DNA extract may be amplified in duplicate, triplicate, or even quadruplicate and repeating alleles may be assigned to the consensus or composite profile. These approaches ensure the accuracy of allelic assignments by identifying allelic drop-ins and drop-outs, and increase the number of determined alleles and loci in the sample (6, 7, 24–26). The described procedure herein is based on the independent amplification of three aliquots per sample.

3.1.3. Two Step Analysis

Profile determination goes beyond the generation of a consensus or composite profile. Rather analysis of LT-DNA samples is a two step process. In the first step, confirmed alleles are assigned to the consensus profile. To be confirmed, alleles must be labeled in at least two of the three amplifications. Sample interpretation always has to start with a decision as to whether or not there are enough components in a profile in order to proceed. The interpretation

should stop if, for example, there are fewer than eight repeating alleles over at least six autosomal loci in the consensus profile. Caution should be exercised when there are more than six repeating alleles at a minimum of two loci in the consensus profile.

The second step is to apply the interpretation protocols which are specific to the sample class (mixture vs. single source).

3.1.4. Possible Outcomes   A sample can be interpreted as a "single source" (when it is clearly possible to assign alleles), a "deconvoluted mixture sample" (when the major donor is determined or assigned), a "mixture for comparison only" (when no major component could be determined), or "inconclusive" (when there are either too few or too many alleles to draw any conclusion).

3.1.5. Interpretation Guidelines for Single Source and Deconvoluted Mixtures   If a locus cannot be deduced according to the interpretation guidelines below, the locus is deemed inconclusive.

3.1.6. Single Source   Special attention must be paid when assigning heterozygotes and homozygotes because of random PCR effects. With HT-DNA samples, often stochastic peak height thresholds are established. However, due to PCR and electrophoresis enhancements, peak heights may be tall for even extremely low amounts of DNA (Fig. 1).
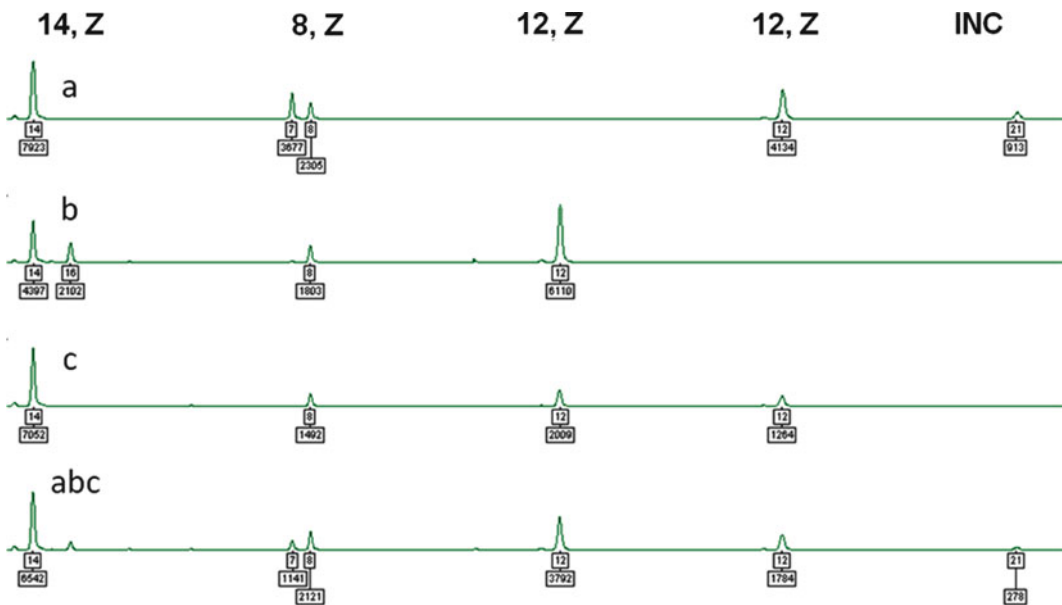
Fig. 1. Interpretation guidelines, not stochastic peak height thresholds, are applied to assign a homozygote in a single source sample amplified with 16.4 pg of template DNA in each of three replicates. A homozygote cannot be assigned at any of the loci shown since one (major) allele is either not present in all three amplifications, or all other alleles are >30% of its intensity. Note that the potential false homozygote allele reaches heights above 7,000 RFUs.

Table 1
The locus is deemed inconclusive when two alleles in one of the three amplifications are completely different compared to the other two amplifications

Replicate "a"	12, 13
Replicate "b"	12, 13
Replicate "c"	14, 15
Consensus or composite profile	12, 13
Assigned alleles	Inconclusive

This precludes the use of a stochastic peak height threshold and necessitates the use of a different set of interpretation protocols.

A heterozygote is determined based on repeating alleles. In general, the two alleles must be the tallest peaks at a locus in at least two out of three amplifications (no peak height ratio requirement). However, there are defined exceptions. If two repeating alleles are clearly major alleles, any additional repeating alleles which are consistently minor, are not assigned. If there is the same repeating allele in the ±4 bp stutter position of another peak, and its intensity in RFU is <30% of the major allele in two out of three amplifications, and, in addition, its intensity is <50% of the major peak intensity in the third amplification, it cannot be a part of the heterozygote pair. In those cases assign the tallest peak and an indicator, such as a Z, that another allele *could* be present.

Homozygosity (not based on RFU threshold), is assigned if the same major allele is present in all three amplifications, and all other alleles (if applicable) are <30% of its intensity (Fig. 1). If an allele, even if it does not repeat, is >30% of the repeating major allele, it could be an indicator of allelic drop-out. The possibility of a heterozygote should be considered. In addition, homozygosity should always be questioned at the following less efficient or higher molecular weight loci (CSF1PO, TH01, D16S539, D2S1338, D18S51, and FGA). In degraded samples, the largest apparent locus should also be considered a possible false homozygote, as well as all loci in samples with less than 20 pg of input template DNA per replicate.

In any case, when the two alleles in one of three amplifications are *completely different* compared to the other two amplifications, the locus should be deemed inconclusive (Table 1).

3.1.7. Mixtures   If there are at least three repeating alleles at a minimum of three loci (3 × 3 rule) a sample may be considered a mixture. Even if there are not at least three loci with at least three repeating alleles, if the results of different replicates are not consistent with each other, a sample should be considered a mixture. Once a mixture
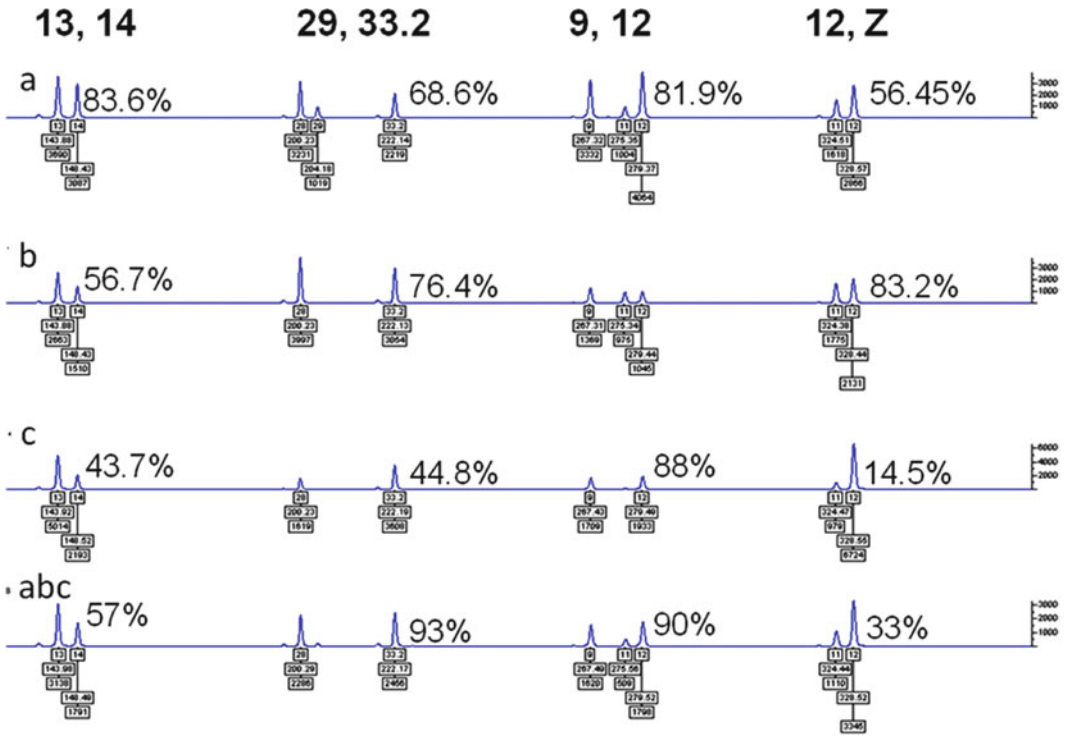
Fig. 2. Deconvolution of a mixture amplified with 50 pg of template in each replicate. Alleles are assigned to the major component if the alleles are labeled in all three replicates and they are major peaks in at least two replicates. Moreover, the peak height ratio of the heterozygote pair must be at least 50% in at least two amplifications. Note that in the "b" replicate of the third locus, the peak heights are too close to ascertain which allele is the tallest. At the fourth locus, the peak height ratio of the 11 allele to the 12 allele in the pooled "abc" sample, does not confirm the 11, 12 assignment from the individual replicates, and therefore the pair cannot be deduced.

has been established, the most important and, at the same time most challenging task in LT-DNA interpretation, is to deconvolute the mixture, or assign a DNA profile to the major contributor.

A mixture should only be deconvoluted if the ratio of the contributors is greater than 2:1. The pooled amplification product is helpful in this determination, due to the fact that heterozygote pairs are generally more balanced in the pooled sample than in the individual replicates. The validated rules are: Alleles can be assigned to a major component if they are labeled in all three replicates and if they are major peaks in at least two replicates (Fig. 2). A heterozygote duo is assigned if the balance between the two peaks is at least 50% in two out of the three amplifications. Homozygosity in mixtures has to be deduced with extreme caution. If an allele is major in all three amplifications, and all other peaks are <30% of its intensity, it can be assigned as a homozygote. In all other cases, assign an indicator of the possible presence of another allele. There are several instances where heterozygosity should always be suspected. These are the larger, higher molecular weight and the less efficient loci

(CSF1PO, TH01, D16S539, D2S1338, D18S51, and FGA) and TPOX. Prone to primer binding site mutations, TPOX is especially concerning when components of mixture share the same alleles (27). False homozygosity should also be suspected and an indicator assigned at the last apparent locus in each color and all loci for samples with input template DNA less than 20 pg per replicate.

For mixture interpretation, the pooled sample (independent CE run of the combined amplified replicates) can be useful although it must not be analyzed independently. Rather, caution should be exercised if the pooled sample does not confirm allelic assignments from the triplicate amplifications (Fig. 2).

The deduction guidelines above apply to major components only. If a contributor is assumed, the DNA profiles of the minor component may be assigned based on the repeating alleles that cannot be attributed to the assumed contributor.

3.1.8. Mixtures (or Mixture Components) for Comparison Only

Samples that cannot be deconvoluted may be used for comparison with a known profile. Results from all amplifications are utilized including nonrepeating alleles, since they could come from the real (minor) source of DNA. However, for qualitative comparisons, only loci with at least one repeating allele are considered. Several scenarios based on qualitative comparisons are possible: "Included"—a person could be a contributor to the sample if all alleles are labeled in the results.; "Cannot Be Excluded"—if one or two alleles consistent with those of a known sample are not apparent in a mixture (or below threshold); "No Conclusions"—no determination can be made since, for example, three alleles consistent with those of a known samples are not apparent in a mixture at less efficient loci; "Excluded"—if there is insufficient evidence to support that a known sample could have contributed to a mixture.

"Cannot Be Excluded" Statement—Qualitative statements provide a valuable piece of information to the court, but must be explained adequately. Specifically, the parameters that define the use of the statement should be clearly stated. There are scientific explanations for the absence of the allele(s) consistent with the known sample. This absence can be attributed to the amount of DNA amplified, artifacts such as stutter, degradation, empirically defined locus characteristics, the length of the STR repeat, and/or the number of contributors to the sample.

In order to assign a statistical value to these types of comparisons, statistical software has been developed. This software calculates the likelihood ratio (LR) of the probability of a given mixture if a particular individual, such as a suspect, did or did not contribute to the mixture (NYC OCME, manuscript in preparation). For these quantitative comparisons, results from all amplifications are utilized, including loci with no repeating alleles. The software developed and validated at the NYC OCME, the Forensic Statistical Tool (FST), employs empirically determined drop-out and drop-in rates based on single source samples and mixtures amplified with a

range of template DNA amounts, mixture ratios, and the number of contributors. Other methods employ probabilistic estimates of drop-out and drop-in rates (28–31).

3.1.9. Technical Review

Mixture identification and deconvolution is certainly a challenging process. Interpretation protocols, training, and supervisory review serve to resolve this issue. It is highly recommended to have multiple levels of review: initial analysis and technical review of the sample and control electrophoretic results; analysis and technical review of case results, and for complex mixture deconvolutions, an additional level of technical review.

3.2. Courtroom Specifics

The relevance of the absence or the presence of someone's DNA in a mixture depends upon the context of the case. Forensic analysts are only able to report and explain the results of the DNA they are able to detect in a sample. They cannot determine when the sample was deposited, how long the sample was there, or how the sample was deposited. Nevertheless, analysts can relay results from studies that demonstrate that with respect to "touch" DNA, it is highly unlikely that one could attribute a DNA profile in a case to secondary or tertiary transfer (32).

In conclusion, following internally validated LT-DNA quality control, testing and interpretation methods consistently yield reliable results. When establishing these protocols, laboratories must demonstrate through their own validation procedures that their methods, which might include any adjustments, are robust and reproducible. To implement these procedures, analysts and technical reviewers must be properly trained to interpret and explain the results and their weight.

4. Notes

1. Facility conditions.
   - Proper protective attire: full laboratory gowns, face masks covering nose and mouth, gloves (double gloves) for bench work and evidence examination, hair coverings, booties if available.
   - Testing performed in dedicated, protected workspaces: Perform all bench work and examination of LT-DNA items under a hood. Each process is performed under a different hood space if available.
   - Decontamination of equipment: Surface decontamination of pipettes, cap openers, racks, hoods, counters, centrifuges, robots, etc.—Rinse several times with 10% bleach stored in the specialized containers that mix the bleach contents as they are used. Rinse with 70% Ethanol. Rinse

with distilled water. Irradiate or fume consumables such as plates and tubes, swabs, and water (see below).

- Daily maintenance: Clean work area before and after each operation. For hoods, apply UV light for 10 min after use.

- Weekly and monthly laboratory clean-up: Wipe down all benches, hoods, instruments, pipettes, cap openers, scissors, and tweezers with bleach, water, and ethanol. Replenish laboratory supplies as needed according to the inventory lists. Irradiate or fume consumables such as plates and tubes, swabs, and water (see below).

- Irradiation in a Stratalinker® 2400: UV irradiation "breaks down" DNA so that DNA testing is not compromised. All plasticware, water, and buffers are irradiated in a Stratalinker® UV Crosslinker 2400 (Stratagene, La Jolla, CA, USA) prior to use. The strength of the UV light in the Stratalinker® should be tested weekly.

- Other preventive measures: Two sets of disposable gloves should be worn in order that the outer gloves may be changed frequently without exposing the skin. Aerosol-resistant pipette tips should be used and changed on every new sample to prevent cross-contamination during liquid transfers. For manual processes, only one sample tube should be opened at a time, and tubes should be opened with a cap opener and/or clean tissue. For all practices, tissue culture technique should be employed such that there is no crossing over of open tubes or reagents.

2. Amplification kits should be tested for sensitivity down to, for example, 6.25 pg. For this test, the sample must generate the expected number of determined alleles using the interpretation protocols. Kits should be retested periodically.

3. Routine temperature monitoring on thermocyclers and QRT-PCR instruments and pipette calibration should be conducted. To ensure that the capillary electrophoresis (CE) machines are functioning properly, the height of a size standard peak in amplification negative controls should be monitored within a determined interval.

4. According to the validation studies of the protocols presented here, 89% of the time the controls were clean. The detected alleles were random and not due to gross contamination. The threshold for the total number of spurious alleles tolerated over three negative control amplification replicates was determined by adding the average number of labeled alleles observed in three amplifications replicates that displayed drop-ins plus two standard deviations of the average. If more alleles are detected, the batch should fail. If the sensitivity of the system increases,

the number of total drop-ins detected over three replicates may increase and accordingly the threshold may increase.

However, as an additional precaution, if a labeled allele repeats in two of three replicates of a negative control, any sample having that labeled allele at that locus should be deemed inconclusive for that locus. One or two repeating alleles are tolerated since one cannot attribute the source of such a small amount of DNA. However, if more than two alleles repeat over three replicates of a negative control, the batch fails. In fact, the number of repeating alleles permitted in negative controls should not increase as this indicates confirmed DNA.

5. At the New York City Office of Chief Medical Examiner samples are sampled with a patent-pending swab premoistened with 0.01% sodium dodecyl sulfate (SDS). Swabbing is performed with a light touch and, if applicable, with the grain of the item. If needed, more than one swab per item or section of an item is used. Due to the use of detergent, DNA is extracted from swabs as soon as possible following collection.

LT-DNA extraction: Samples are incubated in 0.05% SDS and 0.72 mg/mL proteinase K at 56°C for 30 min with shaking at 1,400 rpm, then at 99°C for 10 min without shaking. Following centrifugation, the digest is purified and concentrated twice with Microcon® 100s (Millipore, Billerica, MA, USA) pretreated with 1 µg of fish sperm to prevent DNA loss. The DNA is eluted with 20 µL of water, and is measured and amplified within 1–3 days of extraction.

6. Two microliters of sample is quantitated on the Rotor-Gene Q 3000® (Qiagen, Valencia, CA, USA) using an Alu-based real time PCR assay based on the method described by Nicklas and Buel ([33]), with the exception of the addition of 0.3 µL of 100X SYBR green I (Molecular Probes) and 0.525 mg/mL BSA in a 25 µL reaction volume. The concentrations of standards and calibrators should be verified in triplicate three times, for a total of nine tests using a different mode of quantitation (such as a spectrophotometer). In order to prevent DNA loss, these standards and calibrators should be stored in single use aliquots at –20°C in high concentrations. The dynamic range of the assay extends from 0.39 pg/µL to 1600 pg/µL and measurements are within 30% of their expected value. The no template control threshold was set at 0.1 pg/µL.

7. Amplification protocol: Soak at 95°C for 11 min; 31 Cycles (Denature at 94°C for 1 min, Anneal at 59°C for 2 min, Extend at 72°C for 1 min); 60 min incubation at 60°C; Storage soak indefinitely at 4°C.

8. CE injection parameters: Samples may be injected at either 1 kV for 22 s (low), 3 kV for 20 s (normal), or 6 kV for 30 s

(high) on the ABI Prism® 3100xl Genetic Analyzer (Applied Biosystems). All 150, 100, and 75 pg samples are injected low; 50, 25, and 20 pg samples are injected normal; and 12.5 and 6.25 pg samples are injected high. For the low and the normal injections, 0.5 µL of allelic ladder and 1 µL of a 1/10 dilution of the positive control (100 pg template) with 0.375 µL of LIZ are prepared with HIDI formamide in a total volume of 16 µL. For the high injection, 0.3 µL of allelic ladder and 1 µL of a 1/20 dilution of a positive control (100 pg template) are prepared in the same fashion. Data are collected using nonvariable binning and a base line window of 250.

References

1. van Oorschot, R. A., Ballantyne, K. N., and Mitchell, R. J. (2010) Forensic trace DNA: a review, *Investig Genet 1*, 14.

2. Capelli, C., Tschentscher, F., and Pascali, V. L. (2003) "Ancient" protocols for the crime scene? Similarities and differences between forensic genetics and ancient DNA analysis, *Forensic Sci Int 131*, 59–64.

3. Findlay, I., Ray, P., Quirke, P., Rutherford, A., and Lilford, R. (1995) Allelic drop-out and preferential amplification in single cells and human blastomeres: implications for preimplantation diagnosis of sex and cystic fibrosis, *Hum Reprod 10*, 1609–1618.

4. Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L. P., and Bouvet, J. (1996) Reliable genotyping of samples with very low DNA quantities using PCR, *Nucleic Acids Res 24*, 3189–3194.

5. Findlay, I., Taylor, A., Quirke, P., Frazier, R., and Urquhart, A. (1997) DNA fingerprinting from single cells, *Nature 389*, 555–556.

6. Gill, P. (2001) Application of low copy number DNA profiling, *Croat Med J 42*, 229–232.

7. Caragine, T., Mikulasovich, R., Tamariz, J., Bajda, E., Sebestyen, J., Baum, H., and Prinz, M. (2009) Validation of testing and interpretation protocols for low template DNA samples using AmpFlSTR Identifiler, *Croat Med J 50*, 250–267.

8. Buckleton, J. (2009) Validation issues around DNA typing of low level DNA, *Forensic Sci Int Genet 3*, 255–260.

9. Petricevic, S., Whitaker, J., Buckleton, J., Vintiner, S., Patel, J., Simon, P., Ferraby, H., Hermiz, W., and Russell, A. (2010) Validation and development of interpretation guidelines for low copy number (LCN) DNA profiling in New Zealand using the AmpFlSTR SGM Plus multiplex, *Forensic Sci Int Genet 4*, 305–310.

10. Whitaker, J. P., Cotton, E. A., and Gill, P. (2001) A comparison of the characteristics of profiles produced with the AMPFlSTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis, *Forensic Sci Int 123*, 215–223.

11. Strom, C. M., and Rechitsky, S. (1998) Use of nested PCR to identify charred human remains and minute amounts of blood, *J Forensic Sci 43*, 696–700.

12. Westen, A. A., Nagel, J. H., Benschop, C. C., Weiler, N. E., de Jong, B. J., and Sijen, T. (2009) Higher capillary electrophoresis injection settings as an efficient approach to increase the sensitivity of STR typing, *J Forensic Sci 54*, 591–598.

13. Kloosterman, A. D., and Kersbergen, P. (2003) Efficacy and limits of genotyping low copy number (LCN) DNA samples by multiplex PCR of STR loci, *J Soc Biol 197*, 351–359.

14. Roeder, A. D., Elsmore, P., Greenhalgh, M., and McDonald, A. (2009) Maximizing DNA profiling success from sub-optimal quantities of DNA: a staged approach, *Forensic Sci Int Genet 3*, 128–137.

15. Smith, P. J., and Ballantyne, J. (2007) Simplified low-copy-number DNA analysis by post-PCR purification, *J Forensic Sci 52*, 820–829.

16. Forster, L., Thomson, J., and Kutranov, S. (2008) Direct comparison of post-28-cycle PCR purification and modified capillary electrophoresis methods with the 34-cycle "low copy number" (LCN) method for analysis of trace forensic DNA samples, *Forensic Sci Int Genet 2*, 318–328.

17. Schiffner, L. A., Bajda, E. J., Prinz, M., Sebestyen, J., Shaler, R., and Caragine, T. A. (2005) Optimization of a simple, automatable extraction method to recover sufficient DNA from low copy number DNA samples for

generation of short tandem repeat profiles, *Croat Med J 46*, 578–586.

18. Gill, P., Whitaker, J., Flaxman, C., Brown, N., and 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.

19. Leclair, B., Sgueglia, J. B., Wojtowicz, P. C., Juston, A. C., Fregeau, C. J., and Fourney, R. M. (2003) STR DNA typing: increased sensitivity and efficient sample consumption using reduced PCR reaction volumes, *J Forensic Sci 48*, 1001–1013.

20. Budowle, B., Eisenberg, A. J., and van Daal, A. (2009) Validity of low copy number typing and applications to forensic science, *Croat Med J 50*, 207–217.

21. Tamariz, J., Voynarovska, K., Prinz, M., and 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.

22. Shaw, K., Sesardic, I., Bristol, N., Ames, C., Dagnall, K., Ellis, C., Whittaker, F., and Daniel, B. (2008) Comparison of the effects of sterilisation techniques on subsequent DNA profiling, *Int J Legal Med 122*, 29–33.

23. Cowen, S., Debenham, P., Dixon, A., Kutranov, S., Thomson, J., and Way, K. (2010) An investigation of the robustness of the consensus method of interpreting low-template DNA profiles, *Forensic Sci Int Genet.* 2011 Nov;5(5): 400–406. Epub 2010 Oct 16.

24. Benschop, C. C., van der Beek, C. P., Meiland, H. C., van Gorp, A. G., Westen, A. A., and Sijen, T. (2010) Low template STR typing: Effect of replicate number and consensus method on genotyping reliability and DNA database search results, *Forensic Sci Int Genet.* 2011 Aug;5(4):316–328. Epub 2010 Jul 22.

25. Balding, D. J., and Buckleton, J. (2009) Interpreting low template DNA profiles, *Forensic Sci Int Genet 4*, 1–10.

26. Gill, P., Puch-Solis, R., and Curran, J. (2009) The low-template-DNA (stochastic) threshold -its determination relative to risk analysis for national DNA databases, *Forensic Sci Int Genet 3*, 104–111.

27. Butler, J. M., Shen, Y., and McCord, B. R. (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA, *J Forensic Sci 48*, 1054–1064.

28. Tvedebrink, T., Eriksen, P. S., Mogensen, H. S., and Morling, N. (2009) Estimating the probability of allelic drop-out of STR alleles in forensic genetics, *Forensic Sci Int Genet 3*, 222–226.

29. Gill, P., Curran, J., and Elliot, K. (2005) A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci, *Nucleic Acids Res 33*, 632–643.

30. Evett, I. W., Gill, P. D., Jackson, G., Whitaker, J., and Champod, C. (2002) Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks, *J Forensic Sci 47*, 520–530.

31. Buckleton, J. S., Curran, J. M., and Gill, P. (2007) Towards understanding the effect of uncertainty in the number of contributors to DNA stains, *Forensic Sci Int Genet 1*, 20–28.

32. Tamariz, J., Wong, A., Perez, J., Mitchell, A., Prinz, M., and Caragine, T. (2009) Investigation of the detection of DNA through secondary and tertiary transfer, In *20th International Conference on Human Identification, Las Vegas, Nevada. October 12–15, 2009.*

33. Nicklas, J. A., and Buel, E. (2003) Development of an Alu-based, real-time PCR method for quantitation of human DNA in forensic samples, *J Forensic Sci 48*, 936–944.

Chapter 15

Interpretation Guidelines for Mixed-STR Multilocus Electrophoretic Profiles

Juan Antonio Luque

Abstract

The interpretation of multilocus STR profiling is a complex task that requires a high degree of expertise due to the high number of variables that can affect a biological and analytical process like this.

The purpose of this chapter is to provide guidelines to categorize the mixtures using a discrete system and select those in the white area to produce a profile for reporting and/or comparing.

Key words: DNA analysis, Mixtures, LR, STRs, Interpretation guidelines

1. Introduction

The interpretation of multilocus STR profiling is a complex task that requires a high degree of expertise due to the high number of variables that can affect a biological and analytical process like this. If we analyze the same sample twice, the results will not be fully identical. When we analyze mixed profiles, the degree of complexity increases, and we must take into account additional factors like the number of individuals in the sample, the ratio between profiles, allele masking, and the fact that mixture of two profiles is not 100% additive (see Note 1).

There are two main approaches for the analysis of DNA mixed and single-source profiles (each one with its pros and cons):

A discrete system based on thresholds that determine a valid profile/allele when it is above those thresholds. It consists of three steps: profile determination, comparison with reference samples, and statistical evaluation (RMNE or LR) in case of compatibility/match. Unfortunately, there are a number of situations around the threshold difficult to solve, where the experience of the expert is crucial.

- A continuous system based on probabilities which in one step compares any peak from the evidence with reference samples, obtaining a final value of LR conditioned to some events like allelic drop-out, stutters, heterozygous imbalance, … each one with an associated probability.

The robustness of a DNA profile can be classified in a gradient from 0, where we would have no certainty (baseline level profile), to 1, where we would have absolute certainty of the result (utopia). Some authors (1) have defined three categories for single-source profiles symbolized with three colors: red, amber, and green. The red zone comprises low template DNA (LTDNA) profiles where extreme drop-out phenomena are common, while the green area comprises "conventional" profiles where the possibility of drop-out is negligible. There will be an intermediate amber zone. In the case of mixtures, the situation is more complex, but the categories can be defined similarly. We prefer to symbolize the categories with a black area where the results are not conclusive, a white area where, with a proper routine, we can have a good degree of confidence in the results, and finally a gray area in between where LTDNA components or complex combinations make the interpretation difficult. In mixed profiles, the limits are not defined, so a good analytical phase and a high degree of expertise in the interpretation can reduce the gray area in favor of the white area (at least in some components).

In the continuous system, there are no categories. The black side would be the equivalent to 0 (bottom of the black area in the discrete system) and the white side would be the equivalent to 1 (top of the white area in the discrete system) with a continuous gradient of gray in between; the closer it is to the black side, the more intense the gray is, depending on the probability value. We will have a series of "good" profiles, using the discrete system (white area) or using the continuous system (area next to white where the probability of drop-out and other phenomena is negligible). There will be another series of profiles unusable (black area—discrete system—and area next to black-continuous system). In the intermediate zone (gray) with the continuous system, the value of LR will reflect the degree of certainty. In the discrete system, the results in this gray area can be considered as inconclusive. An experienced analyst can move many profiles to the white area, but it is a subjective system and may be subject to errors. Therefore, if we fix high thresholds (very safe), we can leave many unsolved cases, and if we are more permissive, we solve more cases, but sometimes wrongly (especially if you do not have vast experience). It seems clear that continuous system is more advantageous, however not without drawbacks. On the one hand it is complex to implement and it requires expert systems, or at least some computer tools that you have to train (as an analyst) and to validate intensively. On the other hand, you need to incorporate an LTDNA

framework. Finally, the continuous model does not give a profile for entry into a national criminal database, for exchange between laboratories or simply to compare results from different analysts.

Therefore, the best solution is a combined system where there is a series of steps to categorize the profiles, and to evaluate directly those in the white area, while the profiles of the gray and black area require LTDNA strategies and expert systems based on the continuous method, to help the analyst to take decisions and report results in numerical form.

The purpose of this chapter is to provide guidelines to categorize the mixtures using a discrete system and select those in the white area to produce a profile for reporting and/or comparing. We do not try to make a comprehensive compilation of all possible situations that may arise in the interpretation of mixtures, because the large number of variables and situations would far exceed the scope of this chapter. Instead we opted for a series of steps that forces us to evaluate most of the issues and prevent to make mistakes (maximizing the white area with a reasonable safety). In many cases, we have questions rather than answers, since many aspects are ambiguous and cannot be resolved, or there is no consensus on how to solve them, but at least we can rule out incorrect answers. Some aspects of the interpretation of mixtures as the criteria for analysis and interpretation of LTDNA will be detailed in other chapters. We refer to them. Regarding the continuous system, since it depends on expert systems, it requires no guides, but a validation process. We refer to specific publications on these systems.

2. Methods: Interpretation Guidelines for Mixed-STR Profiles

The first step before interpreting mixed profiles is to have a quality profile. The analyses of single-source samples (not mixed) are less stringent about the profile quality than the analyses of mixtures. The laboratory's quality requirements should be high in case of dealing with mixtures (see Note 2).

Do not try to interpret mixed profiles of poor quality, best, try to improve the analysis before making an interpretation of mixtures (see Note 3). If you have some loci fully drop-out, we must be aware that there may be additional allelic losses, degradation, and stochastic effects. Where possible, the evaluation of incomplete mixed-profiles should be avoided. It must always evaluate the full profile, not locus-by-locus independently. Most parameters are from the whole profile (e.g., mix ratio or number of contributors). It is not appropriate to use different assumptions for each locus.

2.1. Call Peaks and Alleles	The first step is to analyze the samples and controls with the editing program. It is important, especially when working with mixtures, that the control sample results should be correct. Poor quality results in control samples are indicative of that part of the process was not correct, and therefore the results may be affected and not follow the parameters set in the validation of the method. In an EPG we can find:

- Allelic peaks from sample.
- Nonallelic peaks produced by artifacts in extraction/amplification (mostly stutter and nonspecific amplification) or in the detection system and software (mainly spikes and pull-up).
- No peaks. Those constituents of the sample that do not appear in the profile. It is important to determine in mixed samples when it can arise drop-out (see Note 4).

The more relevant nonallelic peaks in mixtures are the stutters. If there are components of the mixture at stutters level, these must be evaluated.

The software must be configured to check all peaks above the limit of detection (LOD) (see Note 5). It is important to turn off automatic stutter removal in software, so that all peaks are recognized. At a later point it will be evaluated if those are stutters or possible alleles.

All peaks detected are recorded for the following steps (see Notes 6 and 7).

2.2. Characterize Mix	Having identified the possible constituents of the mixture, it must be characterized by a number of parameters to use in subsequent steps. At this point, the determination of these parameters need not be accurate because at a later step we will refine the profile. Where statistical evaluation will be performed, we will reevaluate the parameters with the refined profile. The determination of the parameters is done recursively, so a change in a parameter will change another (e.g., a change in the number of contributors needs to recalculate the mix ratio). Similarly, the separation between major and minor profile can change the parameters (see Note 8).

2.2.1. Determine Number of Contributors	This step is the starting point for the entire routine. We determine if we have a single-source profile or a mixture (see Note 9). If you determine that it is a single-source profile, all subsequent steps are avoided, going directly to the comparison with reference samples or the introduction into the database. In mixed samples, we must determine the minimum number of contributors. This number may be distinct from the actual number of contributors. We can estimate the maximum number of contributors that it is reasonably consistent with the results.

As a general rule, while determining the number of contributors, peaks at stutter positions that do not exceed the threshold (ST) are not taken into account.

- If there are two or more loci with three or four alleles, the number of contributors will be established initially as two or more (see Notes 10 and 11).

- If there is a locus with more than four alleles (see Note 11), the number of contributors will be established as three or more (see Note 12).

- If five or more alleles are observed in at least half of the loci of a profile, it should be suspected of more than four individuals. It is recommended to use these profiles only to exclude individuals (see Note 13) unless major/minor components can be separated.

2.2.2. Determine Profile Mix Ratio

The ratio of components is a global property of the mixture, depending on the relative amount of DNA from each component, and therefore must be evaluated locus-by-locus for the whole profile. Several factors such as degradation, stochastic effects, or inhibitors can cause the ratio not observed homogeneous in all loci. Mix ratio can be evaluated only in mixtures of two components, or mixtures of more components if they have a separate major profile of two components. In the latter case, we must take into account that there may be the phenomena of masking and/or overlap between alleles and stutter peaks.

For evaluating the profile mix ratio, the process is as follows:

- We will use the loci showing four alleles.

- If there is no degradation, all loci have a similar signal, and therefore we will use all the loci with four alleles.

- If there is degradation, select those loci with four alleles above T_{RFU}. If all loci show the minor component below T_{RFU}, we will select those loci where the major component is above T_{RFU}. If all alleles at all loci of four alleles were found below T_{RFU}, we will use all of them but the estimate may not be accurate.

- The mix ratio at each locus is the sum of the two higher peak alleles divided by the sum of the two lower peak alleles (see Note 14).

- The profile mix ratio is the average of all loci. If there is any outlier locus, it may be discarded from the calculation.

2.2.3. Determine Mix Heterozygous Balance

Heterozygous balance (really, it would be imbalance) varies between loci due to stochastic and analytical effects. Each laboratory, based on working conditions and status of equipment, will observe more or less heterozygous balance. Each markers system has more or less heterozygous balance (Hb) according its optimization. The conditions and purity of each DNA extract influences Hb.

In single-source samples, the heterozygous balance is the ratio in height between the smallest peak and the highest peak. The maximum allowed for nondegraded samples in excess of 500 pg is >0.6 (2). The value of all loci in mixtures cannot be accurately predicted because of overlapping phenomena, it can only be calculated on loci showing four alleles (for the major component and for the minor component separately). We will use the mixture heterozygous balance (mHb) defined as the average of the heterozygous balances of each component from all loci where four alleles were detected. This parameter is used as an indicator of the quality of the profile. A high heterozygous imbalance makes us think that the quality of the sample is not good and may have a predominance of stochastic phenomena.

2.3. Categorizing Mix

In mixed profiles, each component can be different and one or more components may be LTDNA or be degraded (see Note 15). In these cases, if you can clearly separate profiles, each one can carry a different analysis routine ("conventional" or LTDNA strategy). If you cannot separate them, the values determined in the previous section can vary between loci, applying LTDNA strategy for the entire profile becoming necessary.

The aim of this step is to determine what type of mix we have in terms of the proportion of components and try to separate the major component (locus-by-locus).

2.3.1. Separation of Components

We will use a mnemonic, 3-3-3, to set the criteria (>3:1 profile mix ratio, >3:1 major/minor height ratio, 0.3 imbalance = mHb >0.7). As a general rule, two component mixtures with a ratio less than 3:1 should not be separated because there is a high risk of making erroneous assignment (see Note 16). Also as a general rule, the allele ratio to be separated must be higher than $(n+1):1$, where n is the number of contributors (in mixtures of two individuals to separate a major from a minor allele, it is a must to have a height ratio of 3:1 between alleles). This rule can also be used to discard peaks in stutter position. Finally, the mix heterozygote balance should be high (>0.7) (see Note 17).

For two components mixes over the T_{RFU}, with a *profile mix ratio >3:1,* we can use the following logic (don't use in profile mix ratio <3:1, degraded samples nor LTDNA samples):

- In four alleles loci with two components, there is no problem, the two highest peaks are the major components (see Note 18).
- In two alleles loci if the major/minor ratio is >3:1, the major one can be separated as homozygous. If the proportion of major/minor is less than 2:1, the major component is heterozygous for both alleles (see Note 19).
- In one allele loci, the major component is that allele.

- In three alleles loci, the origin can be two heterozygous sharing one allele or a homozygous plus heterozygous. If the allelic ratio is >3:1, the major component can be separated (homozygous or heterozygous). If the profile is well balanced (Hb >0.8) and the major components are heterozygous, with a ratio >2:1, it is sufficient to separate the major one.

2.3.2. Mixtures Categories　As a result of the previous point, samples can be classified into three categories (3), taking into account the full profile:

- Type A: Mixtures of indistinguishable contributors, where it is not possible to distinguish a single-source major component and all alleles of all loci are above the T_{RFU}. Samples with clear indications of degradation are not included in this section. The comparison and evaluation should be done on the complete profile.

- Type Am: Mixtures with a mixed major component. These are mixtures in which a major profile, composed of two or more individuals, has been separated. In this case, the mixed major profile must be evaluated separately as a mixture of Type A.

- Type B: Mixtures with a major single-source component. Mixtures with a separated major single-source component across all loci. This separated profile can be compared, evaluated, and reported as a single-source profile, indicating that it is a major profile plus a minor component. If you have to evaluate/report the no major profile, it must be done like a profile of type A (the complete profile). Only in some loci, it will be possible to establish with certainty the minority profile, which can be used for exclusion. No minor should be reported separately ever, because the alleles may be masked with the major profile.

- Type C: Mixtures with stochastic effects and/or degradation. These include mixed profiles with one incomplete component. The analyst must state the type C profiles in one of the following categories (see Note 20):
  - Interpretable profile with enough quality to continue the process.
  - Profile useful only to exclude (not conclusive for inclusion). In this category fall the most partial profiles if not confirmed.
  - Uninterpretable profile.

2.4. Refining Mixture　Edition of profiles for the evidence must be independent of the reference samples. The profile must be established before proceeding to the comparison or to be entered into databases. It is therefore necessary to check the result by double editing the results. This double editing can be from the same EPG by another analyst or from different amplification or extraction, depending on the

sample and the laboratory routine. If you make double extraction or amplification, it is not necessary to edit each profile in duplicate (in type C and nonmajor type B mix is recommended to always double editing each replica). If the result is the same for all peaks declared as alleles, you will proceed to the next stage. If there are discrepancies, one analyst, both analysts or a supervisor must conduct new editions/replicas to confirm the result.

At this step, if the profiles are of type A, the stutters would be discarded by difference between alleles. If some stutters (below ST) remain and there are no alleles from the mixture at stutters level, they can be removed.

2.5. Compare Mixed Profiles with Reference Samples

Before comparing with the suspect profile, the results must be compared with samples from known contributors. Depending on the mixture, if the number of contributors or the allelic balance throughout the whole profile is not consistent, it may be necessary to recalculate the parameters and repeat the previous steps. For example, in a locus with four alleles where the known contributor is homozygous, if we have declared the sample coming from two contributors, it should be rephrased as composed of three components.

If the result is a mixture of two contributors including the known contributor, alleles in mixed profile not present in the known contributor profile will be considered mandatory, and all must be present in the profile of the suspect to be considered as a match.

At this point, the result can be compared with reference samples. In case of no suspect, the data can be sent to the criminal database. It is highly recommended for samples that are sent to the database to be analyzed in duplicate from the original sample, if possible. If not possible, you should perform at least a second amplification.

If the match with the suspect is full (taking into account the allelic proportion and balance between peaks), it is reported as compatible with the mixture, always accompanied by the appropriate statistical evaluation.

If minor component alleles are at stutters level and match, it will be reported as compatible, together with the statistical evaluation taking into account the possibility that it may be a stutter. It will be declared inconclusive when the match is based solely on stutters and alleles masked by the major (to report it as compatible, at least three loci must have compatible alleles at no-stutter positions—nor masked).

If the suspect has alleles not present in the sample or there are mandatory alleles in the mixture not present in the suspect, the exclusion must be confirmed with a second extraction from the original samples (evidence and reference sample), if possible, to confirm that there has been no contamination or tubes interchange. In parallel, we must review all the parameters to look for erroneous

inferences, especially in type C or minor type B mixtures. If you change a parameter, you must well document the reason and justification for the change; you should not attempt to change the results determined before the comparison, because it is not correct to modify conditions after comparison with reference samples, unless the error was obvious. If you make any substantial changes, it must be confirmed by a new analysis (see Note 21). If duplicates confirm the results, report as not compatible. If you get a compatible result, it will be reported as compatible and accompanied by the corresponding statistical evaluation (it is mandatory to review and document if the initial error has affected other samples).

If the sample is not available for a new analysis, you can analyze/compare with other samples from the same source to confirm the results.

2.6. Statistical Analysis

If the profiles have been reported as inconclusive or uninterpretable, statistical evaluation should not be done.

Whenever compatibility is declared, it must be accompanied by the appropriate statistical evaluation. We prefer the use of LR with alternative hypotheses, because it allows different responses to different situations and individuals. RMNE calculations is an alternative, but this value is characteristic of the mixture, and therefore its value is equal for all individuals supported.

The way to perform the statistical evaluation is complex in some cases and has many nuances. It is beyond the scope of this chapter, so we refer to the specific recommendations and existing publications (2, 4–7, elsewhere).

In some countries, the prosecution and the defense propose hypotheses, but in other countries the laboratory reports an LR with standard assumptions deduced from the available information. If you do so, these hypotheses must be mutually exclusive, and should be attempted to be symmetrical. Report the possibility of evaluating the results with other hypotheses.

The reports should include the parameters, hypothesis and numerical statistical result for LR, together with a concise explanation of the meaning of this value.

3. Notes

1. If we mix two profiles, the heights of common alleles in both constituents are not exactly the sum of individual peaks. This effect can be easily checked with internal size standard: in a typical electrophoresis, all samples have the same amount of internal size standard; however, when comparing the peak height of size standard in different samples we will see differences from sample to sample. In the electrophoresis reagent blank

(which only have formamide and size standard), the peak heights can be several times higher than in the samples.

2. Some important aspects to consider regarding the quality of the profiles when interpreting samples (not exhaustive):

- Degraded samples can show heterozygous imbalance. Furthermore, the existence of nondegraded and degraded DNA makes the mix ratio and the number of contributors difficult to determine. Two common causes of DNA degradation are poor preservation of samples and extraction systems that degrade the sample.

- Good electrophoresis. If the capillary, the buffer, or the polymer is not in good condition, flattened peaks occur in the higher molecular weight alleles, loosing mix ratio and heterozygous balance. Tip: always use fresh buffer and polymer (max 72 h on the sequencer) and not exceed the number of injections for the capillary recommended by the manufacturer or verified in the validation of the method.

- Always work in the optimal signal range where the response is linear to the amount of DNA. Typically, this range is between 500 and 2,000 RFUs. Above 4,000 RFUs pull-up effect can arise. In mixtures it is difficult to get all the components in this range, so you have to adjust the amplification and electrophoresis parameters depending on the mix and target component. For example, in fingernail trace analysis, it is preferable to overamplify slightly the sample without saturating the major component, to bring the minor component to the optimum level. If we analyze a seminal fraction from differential lysis with a residual component of the victim and a male major component, adjust the signal to the major component although some of the minor components can disappear, unless it is of interest. It is important to note that if we suspect the presence of mixed male–male, female–female, or more than two individuals, it is preferable to overamplify the sample, because the result of quantification is the result of the sum of the components, and therefore there is less of each component. In male–female mix, it is easier to adjust the optimum if we make the quantification of both components.

- Keep equipment well maintained to get a good baseline on electrophoresis. The presence of a high baseline is indicative of problems in electrophoresis, with the reagents, with the software analysis parameters and/or matrix, and it may produce the appearance of unexpected peaks. Typically, a detection limit value of 50 RFUs is employed, but not always the analyses meet this criterion. Each laboratory should establish an empirical limit of detection value in real

analysis conditions and review it periodically, checking that the analysis conditions are maintained.

3. If control samples (ladder, electrophoresis blank, extraction blank, positive control, negative control) do not produce the expected results, analyze the cause and repeat the analysis from the point where it failed, once the problem is solved. Sometimes even when all controls are correct, electrophoresis can fail for a sample. Usually, you can detect it in the internal standard size. Repeat electrophoresis of the sample before discarding the results definitively.

   If the signal does not reach the recommended optimum level, amplify again with more amount of DNA, if possible. If you cannot (usually low template DNA (LTDNA)), there are post-PCR techniques to increase the sensitivity (increasing injection time, raising the amount of amplified DNA in the injection tube, concentration of the amplified sample, etc. …). In this case, the interpretation of results cannot be done with the standard conditions. The laboratory must validate the effects of working in these conditions (and generally it should be the employed criteria of LTDNA).

4. Nonallelic peaks:
   - Spikes: They are peaks produced by brief analytical system failure. Their morphology is narrower than allelic peak and sometimes appears in multiple channels. They are usually not reproducible, so they should disappear when repeating the electrophoresis.

   - Stutters: They are generated in the amplification. They tend to be reproducible when electrophoresis and/or amplification are repeated in the same conditions. These are associated with each allele, and can be $-n$ or $+n$ (-4bp or $+4$bp stutters in tetrameric STR). They are inherent in the system and therefore we can determine a stutter threshold (ST) during method validation. +n peaks have less intensity and they usually are not detected. Each laboratory, each locus has a different stutter threshold, so it is important that each laboratory evaluates their stutter thresholds in the same analytical conditions as those used with the samples. The stutters are indistinguishable from alleles and are always present in greater or lesser extent. The larger the peak the greater the stutter associated. In the case of mixtures, LTDNA level components can be combined with stutters to produce signals above the threshold indicating the presence of an allele or may not to exceed this threshold.

5. There is some controversy over the term LOD. Recently, Budowle et al. have redefined the threshold as PAT (peak

amplitude threshold) defined as "The lowest peak height value for which a laboratory will operationally treat an instrumental response as the detection of DNA fragment rather than simple instrument noise." We still use the term LOD for homogeneity with other laboratory analytical techniques, understanding it like the effective limit of detection obtained in the validation of the technique. Any peak above this value is considering valid.

Additionally, Budowle et al. have redefined the stochastic threshold (T_{RFU} as recommended by the ISFG (2)) as MIT (match interpretation threshold).

6. The record of results is a very relevant issue in the evaluation of mixtures. If such entry is not complete, information will be lost. There are several ways to record the DNA profiles:

   • Original electronic file (raw data): It is not practical because it is not edited.

   • Electronic edited file: They require software to review. Not all software can export edited profiles for exchanging. Not suitable for databases and reports.

   • Hard copy: Requires a manual review whenever you need for comparing. Can be included in reports. Not suitable for databases.

   • Table of alleles: It is the simplest for exchanging results, reports, and databases. It loses all the quantitative information. It is best suited for single-source profiles. In mixing profiles, a decision on the peaks to be recorded is required.

   • Table of alleles with height/peak area: It would be the most complete exchange format. Record is more complex. It requires to report the thresholds to evaluate it. It complicates the operation with databases.

   • Table of qualified peaks: Peaks are categorized. The thresholds are implicit. It would be the optimal profile interchanging format between laboratories and databases for mixtures. It requires qualified laboratories and to redesign databases.

7. We think that the future in exchanging of mix profiles should be either an allele table height/peak area or a table with qualified peaks, because criminal databases are designed for sharing and searching of single-source profiles, but in the case of mixtures they could be improved. For this purpose, we propose a 'qualified nomenclature', taking into account that is a tentative. It needs to be further tested and validated before it can be established as a routine nomenclature (see e.g., in Fig. 1):

   • Record all peaks above the LOD separated by "/".

   • Peaks below stochastic threshold are noted in parentheses.
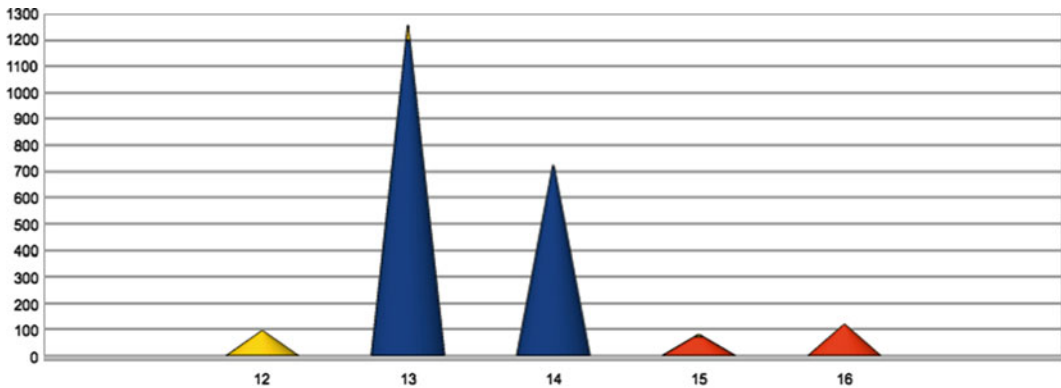
Victim: 15/16

Mix profile: (12s)/13+/14+/(15-)/(16-)

Fig. 1. An Example of peaks annotation in a mixed-STR profile. Record all peaks above the LOD separated by "/." Peaks below stochastic threshold are noted in parentheses. Peaks that do not meet the stutter threshold of an adjacent peak are marked with an "s" after the allele name. Peaks that are in a pull-up position with a peak with DNA excess in another channel are marked with a "p" after the allele name. Alleles consistent with known contributors are marked with "–" after the allele name. Peaks separated as major are marked with "+" after the allele name. In case of suspicion that may have allelic drop-out, a new allele will be introduced: "d" (*lowercase*). If the entire locus is drop-out, a "D" (*uppercase*) will be introduced instead.

- Peaks that do not meet the stutter threshold of an adjacent peak are marked with an "s" after the allele name.
- Peaks that are in a pull-up position with a peak with DNA excess in another channel are marked with a "p" after the allele name.
- Alleles consistent with known contributors are marked with "–" after the allele name.
- Peaks separated as major are marked with "+" after the allele name.
- In case of suspicion that may have allelic drop-out, a new allele will be introduced: "d" (lowercase). If the entire locus is drop-out, a "D" (uppercase) will be introduced instead.

  This approach can be very useful for comparison of sample replicas. Each analyst records the edited analytical profile following qualified notation. A second analyst or reviewer can compare all replicas/editions, debug and/or comparing with reference samples, recording the final reportable profile.

8. For the calculations, there are no significant differences when using the peak area or peak height. Loci with high range of alleles, e.g., FGA, HMW peaks tend to be more flattened (wider and lower) than LMW peaks, being preferably the peak area. On the other hand, if there is partially overlapping alleles,

e.g., 9.3–10 TH01, or "split peaks" problems (–A/+ A), it is usually better to use peak height. For simplicity, we will indicate peak height (the term can be exchanged for peak area).

9. In theory there is never a single-source profile, because always it can appear additional peaks (stutters, drop-in, contamination). In fact, in standard analysis they can't be seen. If using increased sensitivity analysis, these additional peaks can be revealed, so the analysis in these conditions requires special rules (LTDNA/LCN). In practice, once a profile is declared as no apparent mixture, it will be considered for all purposes as a single-source profile. Only if the nature of the case suggests the possible existence of no observed minor components, increase sensitivity analysis will be carried out. For example, in the analysis of traces attached to fingernails, if you perform a conventional analysis based on the amount quantified, a high number of samples in this initial analysis will reveal not possible minor components; a sample overamplification gets the minor components at a detectable level, although most times we are in the range of LTDNA. Today, with real-time quantification of total and male DNA, it is easier to adjust the amount to be amplified in male–female mix.

10. The presence of a single locus profiles with three alleles can be explained by the possible existence of triallelic patterns/ trisomies. In these cases, electrophoresis should be repeated to confirm that it is not a spurious extra peak. If this continues, it must be amplified again, preferably with a different set of primers from those used previously.

11. The loci with four alleles should be examined carefully:
    - If they are not consistent with respect to the mix ratio, we must think that the mixture consists of at least three individuals.
    - If a major and a minor component can be separated clearly, the presence of a major or a minor with three alleles is indicative of a mixture of at least three individuals.

12. If you see more than four alleles, it is difficult to assign a number of contributors. In a simulation with SGM+™ ([8]) 66% of the mixtures of four individuals shows six or fewer alleles. In contrast, for mixtures of three individuals, only 3.3% have four or fewer alleles and for mixtures of two individuals 4.4×10^{-8} have one or two alleles. Thus, the assessment of contributors is quite robust if four or less alleles are observed when taking into account the allelic balance and proportion, whereas not for more alleles.

13. The presence of more than four alleles causes multiple problems. First, we must check thoroughly the absence of contamination

or amplification failure caused by non-specific primers or inadequate temperatures. It is recommended to repeat with another primer pair. Secondly, it is difficult to establish the number of individuals and the relative proportions between them, so a statistical evaluation using the LR is very complex. Finally, the presence of many alleles makes some individuals (with frequent alleles) easily compatible with the mixture.

14. For ratios near 1, may be the selected alleles are from different components due to stochastic fluctuations. This fact does not significantly influence the result.

    In case of extreme mixes, because the minor component presents low signal, estimating the proportion may not be accurate.

15. A very common example in forensic laboratories occurs in vaginal samples. These samples have two components, the epithelial cells of the victim and the aggressor's sperm. Victim's cells in living people are always fresh cells at the time of collection, while the cells of the aggressor (sperm, buccal cells, or other) may be hours or days old, and therefore may have undergone degradation processes. If this situation occurs, usually loci of high molecular weight (HMW) are most affected in male component than in female component, while the low molecular weight loci (LMW) and miniSTRs have less noticeable effects.

16. In some cases cannot be separated both components up to a mix ratio of 8:1, depending on the combination of alleles and the quality of the profile. Usually from a mix ratio of 10:1 the separation of major is safe.

17. If the mHb is high, we can be less restrictive for the difference between alleles. If the mHb is close to 0.6, we need to be more restrictive. If mHb <0.6 it should not try to be separated.
    The values to use can be the following (always with a mix profile ratio >3:1):

    - Imbalance <0.3 (mHb >0.7) apply ratio major/minor >3:1.
    - Imbalance <0.4 (mHb >0.6) apply ratio major/minor >4:1.
    - Imbalance >0.4 (mHb <0.6) did not separate.

18. If you cannot separate (with a profile mix ratio >3:1), a mixture of three or more individuals is most likely involved. In this case, if there is one allele with a proportion to the other three alleles in height >4:1 or <1:3, it can be separated.

19. If the ratio of major/minor is between 3:1 and 2:1, there may be doubts between a mixture of two homozygous and a major heterozygous/minor homozygous mixture.

20. In the Spanish judicial system (guarantor), the principle in dubio pro reo (in case of doubt favor the accused) is used. Our laboratory, as a laboratory of the judiciary system, must give consistent results to the prosecution, and also to the falsely accused person an opportunity to be exonerated. If thresholds are set too high, it will result in excessive inconclusive results, decreasing the chances of defense, and therefore our system favors the prosecution. Therefore, one must assume some degree of risk in setting thresholds for maximum results at least for exclusion purposes. In contrast, if partial results or low statistical weight (complex mixtures) are reported, may be misunderstood by the jury or court, overestimating the statistical significance or assume that not excluded is synonymous with being source from the sample absolutely (DNA testing tends to be overestimated). It is therefore very important in every case where no exclusion or compatibility is reported to include a statistical evaluation, and make clear the significance of that value, and even we can indicate if the statistical significance is low. (Some questions to think about: when is low signification? There is consensus or rationale answer?. Do all juries, judges, prosecutors, or advocates, without a mathematical and scientific background (and every analyst) really understand the significance of an LR or RMNE?).

21. The most common errors that can occur in interpretation (especially if the heterozygous balance is not good or in degraded samples) is the misallocation of one or more peaks to a major/minor component, identify the wrong number of contributors, or eliminate stutters improperly. At the analytical level, tube changes, contamination between samples or from the analyst, or stochastic events (that produce gain/loss of spurious alleles or changes in the proportions) may occur. We must distinguish two situations: the discrepancy is within a few loci or balance of peaks, or the discrepancy will affect many loci. In the latter case if there is no analytical error or contamination, repeating testing will confirm the results as exclusive. If the discrepancy affects few markers, repetition can reproduce the same effects, so it is recommended that the repetition will be done with another set of primers. If the discrepancy is in one marker, you can do complementary analysis to improve results.

References

1. Gill P, Puch-Solis R, Curran J (2009) The low-template-DNA (stochastic) threshold—Its determination relative to risk analysis for national DNA databases. Forensic Sci Int Genetics 3:104–111

2. Gill P, Brenner CH, Buckleton JS et al (2006) DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. Forensic Sci Int 160:90–101

3. Schneider PM, Fimmers R, Keil W et al (2009) The German Stain Commission: recommendations for the interpretation of mixed stains. Int. J. Legal Med. 123:1–5

4. SWGDAM (2010). SWGDAM interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories. Available at http://www.fbi.gov/about-us/lab/codis/swgdam.pdf/at_download/file

5. Evett IW, Weir BS (1998) Interpreting DNA evidence: statistical genetics for forensic science. Sinauer, Sunderland

6. Buckleton J, Triggs CM, Walsh SJ (2005) Forensic DNA evidence interpretation. CRC, London

7. Weir BS, Triggs CM, Starling L et al (1997) Interpreting DNA mixtures. J. Forensic Sci. 42:213–222

8. Buckleton JS, Curran JM, Gill P (2007) Towards understanding the effect of uncertainty in the number of contributors to DNA stains. Forensic Sci Int Genetics 1:20–28

Chapter 16

Capillary Electrophoresis Analysis
of a 9-plex STR Assay for Canine Genotyping

Barbara van Asch and António Amorim

Abstract

STR analysis of canine-derived biological evidence for the identification of individuals is becoming an important tool for forensic investigations. A protocol for the multiplex PCR amplification and capillary electrophoresis of nine autosomal STRs and two fixed-size markers for sex identification in dogs and wolves is described here. The selection of the loci included in the multiplex complies with the recommendations of the International Society for Forensic Genetics in regard to human DNA analysis. The protocol is optimized for automatic fragment size detection in an ABI platform.

Key words: Canine, Forensic analysis, Genotyping, Multiplex PCR, Short tandem repeat

1. Introduction

Short tandem repeats (STRs) are becoming an important tool for the genetic analysis of biological samples of animal origin in the course of forensic investigations (1, 2). Ubiquitous in human environments, dogs are probably the most relevant species in this regard, as they can be victims, offenders, or contribute to link a suspect to a crime scene.

A large number of canine-specific STRs have been reported in the literature (3); however, only a small fraction was developed for use in forensic analysis (4–9). In that context, we have selected a panel of nine canine-specific STR loci and developed a protocol for multiplex PCR amplification and capillary electrophoresis analysis (10).

The method described here, allows for the simultaneous amplification and analysis of nine autosomal STRs (FH3210, FH3241, FH2004, FH2658, FH4012, REN214L11, FH2010, FH2361, and C38) that can be used for the genetic identification of an

individual and parentage testing in dogs. FH3241, REN214L11, FH2010, and FH4012 loci present a perfect tetranucleotide repeat structure, whereas FH3210, FH2004, FH2658, FH2361, and C38 present complex repeat structures. PCR product sizes are relatively short, ranging from 106 to 347 bp (shortest and longest alleles in the panel, respectively). The STRs included in this multiplex were selected based on the following criteria, in agreement with the recommendations of the International Society for Forensic Genetics (ISFG) for human forensic standards: (a) tetranucleotide repeats; (b) flanking regions without palindromic sequences or long polynucleotide stretches; and (c) separate chromosomal locations, avoiding known coding regions and telomeric and centromeric localizations. The multiplex also includes an additional system for sex identification comprised by two fixed-size fragments, one located in the sex-determining region in chromosome Y and the other in the androgen receptor gene in chromosome X. Moreover, this multiplex can also be used for the typing of wolf samples and allows for the statistical discrimination between dogs and wolves, provided that a population database of reference genotypes is previously constructed (11).

Canine-specific STR analysis by PCR amplification followed by capillary electrophoresis follows the same principles as human STR typing. Therefore, general recommendations concerning samples, equipment, and reagent handling should be followed (12). Recommendations for the analysis of nonhuman DNA should also be consulted (1, 13).

2. Materials

Prepare all solutions at room temperature using ultrapure RNAse-free water. Store PCR reagents at –20°C when not in use. Fluorescently labeled primer solutions are light-sensitive and should be stored in dark. Formamide is potentially hazardous and should be handled with care. Use and store pre- and post-PCR samples, reagents, and instruments separately. Use disposable gloves and aerosol resistant pipette tips to prevent cross-contaminations.

2.1. PCR Reagents

1. 2× QIAGEN Multiplex PCR Master Mix (QIAGEN GmbH, Germany) (see Note 1).
2. Labeled and unlabeled oligonucleotide primers (Applied Biosystems) (see Table 1).
3. RNAse-free water (QIAGEN).

2.2. Capillary Electrophoresis

The analysis of PCR products is optimized for ABI PRISM® PCR System 3130xl (Applied Biosystems). The following material list is recommended by the manufacturer.

Table 1
Loci included in the multiplex: chromosomal location, repeat motif as determined by sequencing, PCR product size range and allele range of the nine STR loci included in the multiplex

Locus	Chromosome	Repeat motif	Size range (bp)	Allele range
FH3210	CFA2	$(AAGA)_n^a$	230–315	14–35.1
FH3241	CFA8	$(TTCT)_n$	250–270	9–14
FH2004	CFA11	$(AAAG)_n^a$	161–257	9–33
FH2658	CFA14	$(GAAA)_n$	106–138	9–17
FH4012	CFA15	$(TTTC)_n$	119–143	14–20
REN214L11	CFA16	$(GAAT)_n$	154–162	6–8
FH2010	CFA24	$(TTCA)_n$	154–170	9–13
FH2361	CFA33	$(TTTC)_n^a$	231–347	12–41
C38	CFA38	$(TTCT)_n^a$	132–217	11–32.1
AR	X	Not applicable	185	Not applicable
Sry	Y	Not applicable	106	Not applicable

Chromosomal location and PCR product size are given for two fixed-size markers for sex identification (AR and Sry)
[a]Basic repeat motif in locus with complex structure

1. MicroAmp® Optical 96-Well Reaction Plate (Applied Biosystems, Foster City, CA).
2. HiDi formamide (Applied Biosystems).
3. GeneScan™LIZ®-500 internal size standard.
4. Four-capillary array 36 cm × 50 μm (Applied Biosystems).
5. Performance Optimizer Polymer 7, POP-7 (Applied Biosystems).
6. GeneMapper® Software v4.0 (Applied Biosystems) for fragment size determination.
7. Allelic ladder composed by sequenced alleles for comparative sample typing.

3. Methods

3.1. Loci Information   The nine autosomal STR loci (designation, chromosomal location, repeat motif, PCR product size range, allele size, and allelic range) and the two fixed-size markers for sex determination are summarized in Table 1 (see Note 2).

<table>
<tr><td>

*3.2. Primer Mix
for Multiplex
Amplification*

</td><td>

A primer mix for a large number of samples should be prepared in advance thus avoiding pipetting errors, reducing pipetting time, and contributing to increase PCR reproducibility. Primer sequences, fluorescent labels, and concentration in the multiplex are given in Table 2.

A graphical representation of the multiplex is presented in Fig. 1.

</td></tr>
</table>

1. Prepare 100 µM primer stock solutions by rehydrating the lyophilized primers with the appropriate volume of ultrapure RNAse-free water. Mix well until the primer is completely dissolved.

2. Produce 100 µL of multiplex primer mix (see Note 3). Dispense 58 µL of ultrapure RNAse-free water in a tube. Add 2 µL of each forward primer and each reverse primer, except for the Sry marker (1 µL of each primer). Mix the solution thoroughly and store in dark at –20°C when not in use (see Table 3).

Table 2
Primer sequences and fluorescent labels for the nine autosomal STRs and the two fixed-size markers for sex determination (AR and Sry) included in the multiplex

Locus	Primer	Primer sequence (5′–3′)	Fluorescent label
FH3210	F	CAAGGACCACGATGAAATGACT	VIC
	R	GCTGGATTCAGGAGCTGTTCA	–
FH3241	F	TCCTTGTTTCCTTCCTCTGG	PET
	R	TTGGGCAAAATCAAAACTCC	–
FH2004	F	GGGGCTTTGTACTGTGACCTAC	FAM
	R	ACAGACTGAGAATGCTGGGTT	–
FH2658	F	GAGCTCTACATCTGAAACTATAC	VIC
	R	GAAGACGGTTTGGCAGTTTCT	–
FH4012	F	GGGAGGGAAGCGATCTTCT	FAM
	R	CGGTTAGCCATTCCCTGAG	–
REN214L11	F	GGCTCTCCATGCTAAGACC	VIC
	R	TGGGTCTAATGGTTTGGGATAG	–
FH2010	F	CTATTAACAATGTCAGACTCTCAG	NED
	R	GAGCATGCATGTACACCAGAA	–
FH2361	F	<u>GCAGGTCAGAGCAGTCAGAA</u>	NED
	R	GAATGTACCAGGCACTATGCA	–
C38	F	ACAAGAGGGGATGCTGAA	PET
	R	TCATGTGTCTGTTGGGCATT	–
AR	F	CAGCAGCAGCAGCAGGAG	VIC
	R	CCAGGCTCTGGAACGCAGG	–
Sry	F	GATCAAAGGCGCAAGATGGCT	FAM
	R	CCAGGCTCTGGAACGCAGG	–
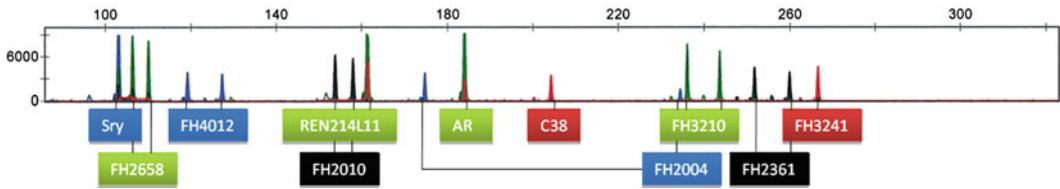
Fig. 1. Genetic profile of a male sample comprising nine autosomal STR loci (FH4012, FH2004, FH2658, REN214L11, FH3210, FH2010, FH2361, C38, and FH3241) and two fixed-size markers for sex identification (AR and Sry).

Table 3
Primer concentrations and volumes for the preparation of the PCR primer mix including 11 loci (nine autosomal STRs and two markers for sex identification)

Primer	Concentration in primer mix (μM)	Volume (μL) for multiplex primer mix (100 μL at 2 μM)	Final concentration in PCR mix (μM)
FH3210 F	2	2	0.2
FH3210 R	2	2	0.2
FH3241 F	2	2	0.2
FH3241 R	2	2	0.2
FH2004 F	2	2	0.2
FH2004 R	2	2	0.2
FH2658 F	2	2	0.2
FH2658 R	2	2	0.2
FH4012 F	2	2	0.2
FH4012 R	2	2	0.2
REN214L11 F	2	2	0.2
REN214L11 R	2	2	0.2
FH2010 F	2	2	0.2
FH2010 R	2	2	0.2
FH2361 F	2	2	0.2
FH2361 R	2	2	0.2
C38 F	2	2	0.2
C38 R	2	2	0.2
AR F	2	2	0.2
AR R	2	2	0.2
Sry F	1	1	0.1
Sry R	1	1	0.1

Primer volumes for the production of the multiplex primer mix are given based on 100 μM primer stock solutions

3.3. Multiplex PCR Amplification

Multiplex PCR amplification is optimized for GeneAmp® PCR System 2700 (Applied Biosystems) thermal cycler. The use of other equipment may require further optimization.

1. Determine the total number of reactions including samples and controls (see Note 4), and add one reaction for each set of 12 samples, in order to compensate for pipetting errors.

2. Thaw reagents at room temperature (if stored at –20°C) and mix the solutions completely to homogenize the concentration of different components in the solution.

3. Dispense appropriate volumes of each reagent in the PCR tubes according to Table 4 (see Note 5).

4. Add the DNA template (see Note 6) to each PCR tube containing the reaction mix and mix gently. Ensure that the tubes are effectively closed.

5. Place the PCR tubes in the thermal cycler using the following program: initial polymerase activation step at 95°C for 15 min, 29 cycles of 94°C for 30 s, 60°C for 1 min 30 s, and 72°C for 1 min, and a final extension at 72°C for 60 min. PCR products can be stored overnight at 4°C or at –20°C for long-term storage.

3.4. Capillary Electrophoresis

The protocol is optimized for use in ABI PRISM® 3130 Genetic Analyzer, under the manufacturer's recommendations. The use of other equipment may require adaptation of this protocol and further optimization. An example genotype obtained using this protocol is presented in Fig. 2.

1. Prepare a loading mixture for all samples to be analyzed by mixing 10 µL HiDi formamide (Applied Biosystems) and 0.25 µL Genescan-LIZ 500 internal size standard (Applied Biosystems) for each sample. Add one reaction for each set of 12 samples, in order to compensate for pipetting errors.

2. Dispense 10.25 µL of the loading mixture into MicroAmp® Optical 96-Well Reaction Plate wells and add 0.5 µL of PCR product.

3. Place the samples in the ABI PRISM® 3130 Genetic Analyzer.

4. Analyze the previously prepared samples using Performance Optimized Polymer-7 (POP-7) and with default Module FragmentAnalysis36_POP7_1 and DyeSet G5. Analyze data using spectral G5 generated using Multi-Capillary DS-33 Matrix Standard Kit (6-FAM™, VIC®, NED™, PET®, and LIZ® Dyes) following the ABI 3130 Genetic Analyzer User's Manual.

5. Determine fragment sizes using GeneMapper® Software v4.0 (see Note 7).

6. Type samples by comparison with the fragment sizes determined for sequenced allelic ladders.

Table 4
Reagent concentrations and volumes necessary for one PCR multiplex reaction

	2× QIAGEN Multiplex PCR Master Mix	Primer mix	DNA (ng)	RNAse-free H$_2$O	Total volume
Concentration in the reaction	1×	0.2 μM	5–10	Variable	10 μL
Volume in the reaction	5 μL	1 μL	Variable	Variable	10 μL
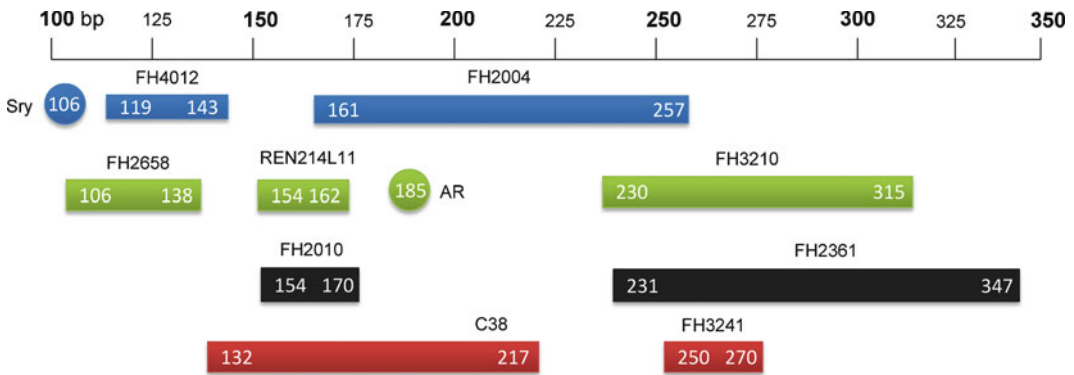
Fig. 2. Schematic representation of observed allelic size ranges and fluorescent primer labels (FAM in *blue*, VIC in *green*, NED in *black* and PET in *red*) for the 9-STR multiplex. The two fixed-size markers for sex identification (Sry and AR) are represented by *circles.*

3.5. Allelic Ladders

The use of sequenced allelic ladders is recommended in order to allow for intra- and interlaboratory comparison of results. An allelic ladder can be constructed by isolating and sequencing different size STR alleles found in a population sample. The DNA samples for which different alleles are identified can be mixed and amplified together to produce a ladder. The volume of each individual DNA sample in the mixture must be adjusted to produce a well-balanced ladder. It is advisable that the ladder for each locus is first produced separately. All ladders can then be combined in a complete multiplex ladder.

1. To produce a DNA mix for the single ladder of each locus, combine approximately equal amounts of each DNA sample (10–20 ng for heterozygous and 5–10 ng for homozygous), aiming at a good balance among alleles (see Note 8).

2. Prepare PCR amplification by combining 1× QIAGEN Multiplex PCR Master Mix (QIAGEN), 0.2 μM of each primer and 3 μL of DNA mix.

3. Run PCR program as follows: 15 min initial activation at 95°C, 32 cycles for 30 s at 94°C, primer annealing at 60°C for 90 s and extension at 72°C for 60 s, and a final extension at 60°C for 80 min.

4. Combine all the individual ladders in a single mixture (multiplex ladder) and assess the amplification balance of alleles by capillary electrophoresis, as described in Subheading 3.4.

5. If adjustment of the relative quantities of each single ladder is necessary to obtain a well-balanced multiplex ladder, repeat from step 5.

6. Dilute the combined multiplex allelic ladder 10^5 times and reamplify for stock, using the same PCR conditions as described above.

4. Notes

1. We routinely use the 2× QIAGEN Multiplex PCR instead of producing a PCR mix from separate components because it greatly facilitates reagent logistics, reduces pipetting effort, and contributes to increased quality and reproducibility of the amplification. The use of other reagents may require further optimization of the PCR amplification protocol.

2. The markers for sex identification consist of: a 106-bp fragment within the coding region of the *Canes lupus familiaris* sex-determining region Y (Sry) gene (Y chromosome, GenBank AF107021, positions 451–556); and a 185-bp fragment within the coding region of the androgen receptor (AR) gene (GenBank AY271347) in the X chromosome between positions 54897911–54898095 (Dog Genome, May 2005 assembly; UCSC Genome Browser Database). These regions were previously described and validated for this purpose (14) but we designed and validated new primers in order to adequately accommodate these markers in the 9-STR multiplex (11).

3. Preparing a volume of 100 μL primer mix for 100 reactions has several advantages: Calculations of the necessary relative volumes of the different primers and H_2O are straightforward; pipetting volumes are not extremely low thus minimizing pipetting errors. This volume also represents a good compromise between a volume sufficiently large for 100 reactions and a relatively low thawing time at room temperature before the preparation of the PCR master mix.

4. We routinely perform negative controls for DNA extraction and PCR amplification. Positive controls are also advisable when working with ancient, degraded, or low copy number DNA samples.

5. PCR amplifications can be performed in volumes lower than 10 μL. In our experience, good results are obtained in volumes as low as 5 μL, provided that relative concentrations of reagents and DNA template are maintained. However, the use of total reaction volumes lower than 5 μL may contribute to low reproducibility due to pipetting errors.

6. Low-purity DNA samples may result in poor amplification and a low signal in capillary electrophoresis. It is advisable that an adequate method of DNA extraction is performed in order to obtain high-purity DNA samples.

7. Degraded and/or vestigial DNA samples may result in reduced or failure of amplification of alleles, particularly in the larger product-size loci such as FH3210, FH2361, and FH3241. The fluorescent label dyes NED™ and PET® also contribute to a decreased signal in capillary electrophoresis.

8. Allelic ladders should include at least the most frequently found alleles. It is advisable to consult the repeat-based standardized nomenclature of these markers, as determined by sequencing (10). Allele frequencies in sample populations are available for dogs (10) and wolves (11).

Acknowledgments

References

1. Linacre A, Gusmão L, Hecht W, Hellmann AP, Mayr WR, Parson W, Prinz M, Schneider PM, Morling N (2010) **ISFG: Recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations.** *Forensic Sci Int Genet.*

2. Amorim A (2010) **Introduction on the special issue on forensic genetics: non-human DNA**. *The Open Forensic Sci J*, **3**.

3. van Asch B, Pereira F (2010) **State-of-the-art and future prospects of canine STR-based genotyping**. *The Open Forensic Sci J*, **3**:45–52.

4. Hellmann AP, Rohleder U, Eichmann C, Pfeiffer I, Parson W, Schleenbecker U (2006) **A proposal for standardization in forensic canine DNA typing: allele nomenclature of six canine-specific STR loci.** *J Forensic Sci*, **51**(2):274–281.

5. Eichmann C, Berger B, Parson W (2004)**A proposed nomenclature for 15 canine-specific polymorphic STR loci for forensic purposes.** *Int J Legal Med*, **118**(5):249–266.

6. Pádár Z, Egyed B, Kontadakis K, Füredi S, Woller J, Zöldág L, Fekete S (2002) **Canine STR analyses in forensic practice. Observation of a possible mutation in a dog hair.** *Int J Legal Med*, **116**(5):286–288.

7. Dayton M, Koskinen MT, Tom BK, Mattila AM, Johnston E, Halverson J, Fantin D, DeNise S, Budowle B, Smith DG *et al* (2009) **Developmental validation of short tandem repeat reagent kit for forensic DNA profiling**

of canine biological material. *Croat Med J*, **50**(3):268–285.

8. Kanthaswamy S, Tom BK, Mattila AM, Johnston E, Dayton M, Kinaga J, Erickson BJ, Halverson J, Fantin D, DeNise S *et al* (2009) **Canine population data generated from a multiplex STR kit for use in forensic casework.** *J Forensic Sci*, **54**(4):829–840.

9. Tom BK, Koskinen MT, Dayton M, Mattila AM, Johnston E, Fantin D, Denise S, Spear T, Smith DG, Satkoski J *et al* (2010) **Development of a nomenclature system for a canine STR multiplex reagent kit**. *J Forensic Sci*, **55**(3):597–604.

10. van Asch B, Alves C, Gusmão L, Pereira V, Pereira F, Amorim A (2009) **A new autosomal STR nineplex for canine identification and parentage testing.** *Electrophoresis*, **30**(2):417–423.

11. van Asch B, Alves C, Santos L, Pinheiro R, Pereira F, Gusmão L, Amorim A (2010) **Genetic profiles and sex identification of found-dead wolves determined by the use of an 11-loci PCR multiplex.** *Forensic Sci Int Genet*, **4**(2):68–72.

12. Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C *et al* (2006) **DNA Commission of the International Society for Forensic Genetics (ISFG): an update of recommendations on the use of Y-STRs in forensic analysis**. *Int J Legal Med*, **120**(4):191–200.

13. Budowle B, Garofano P, Hellman A, Ketchum M, Kanthaswamy S, Parson W, van Haeringen W, Fain S, Broad T (2005) **Recommendations for animal DNA forensic and identity testing.** *Int J Legal Med*, **119**(5):295–302.

14. Eichmann C, Berger B, Steinlechner M, Parson W (2005) **Estimating the probability of identity in a random dog population using 15 highly polymorphic canine STR markers.** *Forensic Sci Int*, **151**(1):37–44.

Chapter 17

Capillary Electrophoresis of DNA from *Cannabis sativa* for Correlation of Samples to Geographic Origin

Heather Miller Coyle

Abstract

For routine genetic analysis of *Cannabis sativa*, two methods are currently in use, (a) AFLP; amplified fragment length polymorphism analysis and (b) STR; short tandem repeat analysis. The AFLP method used on capillary electrophoresis instrumentation is fully described in this chapter. AFLP analysis generates numerous nonspecific marker fragments for a complex DNA pattern and is available in kit format for quality assurance of reagents. This method is particularly useful when discerning the genetics of highly inbred plant species that may share much of the same DNA with only slight differences due to their common genetic background. AFLP analysis, however, is most effective on fresh or well-preserved plant specimens where the integrity of the DNA is high and the sample is a single source specimen (i.e., not a mixture of plants or different species).

Key words: Forensic botany, AFLP, STR, *Cannabis*, Marijuana, DNA, Environmental, Database

---

1. Introduction

In the United States, organized criminal organizations have increasingly used remote areas of American Park Lands for the illegal cultivation of high-potency marijuana. These criminal organizations are (a) threatening the safety of visitors to the Parks, (b) damaging the environment and, (c) funding violent drug cartels (1). These organizations employ gangs and illegal aliens to stay with the illegal crops from planting until harvest and use a combination of threatening devices (booby traps, shotguns, etc.) to keep law enforcement and the unsuspecting tourist from entering the crop area. In addition to public safety, severe damage to the environment is occurring. For every acre of forest planted with marijuana, the equivalents of ten acres are damaged by toxic chemicals (illegal use of fertilizers,

herbicides, etc.). It is estimated that the cost to US taxpayers for restoring the public land to its natural state is approximately $11,000.00 per acre (1). These illegal cultivation sites generate large amounts of standard garbage, biohazard refuse, and toxic waste which seep into the soil and water to affect people, animals, and fish.

Although the local, state, and federal law enforcement agents have taken appropriate action through a coordinated effort by the National Marijuana Initiative (NMI) and funding by Office of National Drug Control Policy (ONDCP) and Department of Justice (DOJ), it is difficult to eliminate the problem since the geographic source (originator) of the material is largely unknown. To aid in geographic sourcing, DNA methods are being employed to characterize the genetics of the *Cannabis* plant and compare against known reference samples as is done with many other forensic databases (2–5).

In forensic science, DNA databases are routinely used for matching evidence to known reference samples to identify a source. This same concept is being developed for DNA matching of seized marijuana from the United States public land back to a series of source samples provided by the Drug Enforcement Agency (DEA) and the NMI to link geographic sources together for intelligence operations and for subsequent prosecution in Federal court. Two methods are currently under consideration: (a) amplified fragment length polymorphism (AFLP) analysis (6, 7) and (b) STR analysis (1, 8–11). Both methods have their own set of benefits and limitations but are relatively simple to perform in the laboratory by experienced molecular biologists. In the biological research at academic institutions, hundreds of different plant species have been successfully genetically analyzed using DNA-based methods (12–17). In this chapter, AFLP analysis will be presented in detail.

AFLP analysis is performed using capillary electrophoresis (CE) instrumentation. Capillary electrophoresis is a standard method for the detection and fine resolution of DNA fragments in an electrically charged field (18). The instrumentation used at our laboratory is the ABI 3130 DNA sequencer and has corresponding ABI Genemapper analysis software. The older standard agarose gel electrophoresis methods employed by many scientists for analysis of polymerase chain reaction (PCR) amplified DNA fragments have been replaced by capillary electrophoresis because CE is more sensitive for DNA detection, limits the use of toxic chemicals for detection, has quality controlled reagents for uniformity in testing, and is semiautomated for faster sample processing. CE uses electro kinetic injection to inject the DNA sample into a fine diameter flexible silicon coated glass tube filled with a gel polymer. An electrical charge is applied and fluorescently tagged DNA fragments will separate based on fragment size and molecular charge. Fragment information is collected by the DNA sequencer and the fluorescent

DNA fragments are imaged by a CCD camera and the collected information (DNA profile) is then available for further analysis.

AFLP analysis uses previously extracted DNA from a single source plant specimen and the scientist creates a DNA pattern characteristic of the sample by using a series of controlled enzymatic reactions to fragment the DNA at specific locations based on sequence (see Fig. 1) and then PCR amplify the fragments and fluorescently label for DNA detection (2, 6, 19, 20) (see Fig. 2). No prior sequence information is required; the reproducibility of the profile is based on the consistent generation of DNA fragments of known size after enzymatic digestion with two different restriction enzymes (e.g., EcoRI and MseI). A variety of other enzymes have been successfully used as modifications to this protocol if the plant species of interest does not exhibit many fragments. This method works well to provide consistent DNA profiles from surface decontaminated seeds and fresh plant leaves, both of which yield high quality DNA (20) (see Notes 1 and 2). Dried plant material and root material that may have bacterial or mold contamination requires that each sample be evaluated prior to AFLP analysis for suitability. This is best done visually and microscopically. The impact for not recognizing contamination is that additional peaks will be present in the final DNA profile that could result in a false exclusion by DNA. In addition, if poor quality (highly fragmented DNA of 500 bases or less) DNA is used with this method; random DNA fragmentation may generate inconsistent AFLP results from sample extraction to sample extraction. To control for this, duplicate extraction may be performed to check for consistency of DNA profiles. If the plant samples are selected carefully and the DNA is of sufficient quality, then AFLP analysis is an excellent method for discerning the genetics of potentially highly inbred plant species

1. DNA extract incubated with EcoRI + MseI restriction enzymes

2. Specific DNA fragments generated by enzymatic cut site recognition

3. DNA fragments with adaptor sequences attached
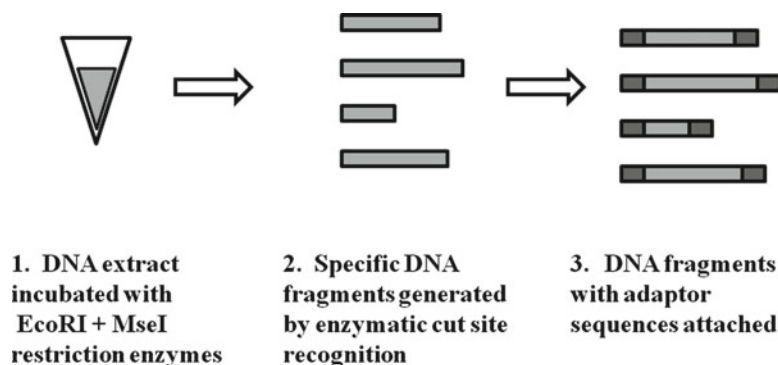
Fig. 1. The initial steps in the AFLP reaction are DNA extraction and digestion using two common restriction enzymes, EcoR1 and Mse1. This enzymatic digestion fragments the DNA at specific sites creating a variety of DNA fragments of different lengths. In order to PCR amplify the fragments, short segments of known DNA, called linkers or adaptors, must be added to the ends of each DNA fragment for PCR primer recognition.
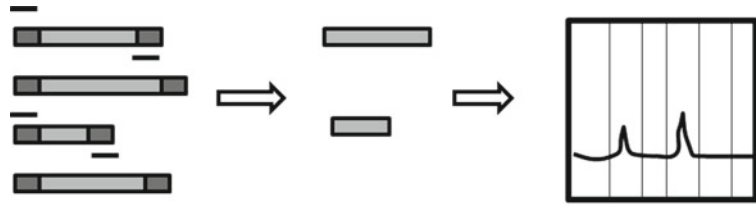
4. **Selective PCR primer pairs amplify a subset of DNA fragments based on sequence recognition to generate a simple scorable pattern**

5. **DNA profile exhibiting 2 peaks; with a binary code for this bin set of 010100**

Fig. 2. Once the linkers have been attached to the fragments, PCR amplification can be performed using a variety of PCR primer combinations to amplify different subsets of DNA fragments based on primer overlap into the DNA fragment. The researcher selects the PCR primer combination that gives the most informative patterns and analyzes using Genemapper software to convert the complex visual peak image to a binary code. The binary code is generated by peak presence or absence whereby presence is scored as a "1" and absence is scored by a "0." The binary code is then the genotype for that sample and can be easily data-based into a variety of searchable software programs.

(see Note 3). AFLP analysis has been successfully utilized on many different crop species including rice (19), corn (19), grapes (12), hops (13), sorghum (14), banana (15), and poplar (16, 17).

2. Materials

All materials and solutions should be prepared with molecular biology grade (DNase and RNase-free) ultrapure water. All reagents should be stored at –20°F (i.e., frozen; kit components will last for up to 1 year) unless in routine day-to-day usage. If reagents are being used every day, they may be stored at 4°F (i.e., refrigeration) with little effect on expiration date as the kit would be completely used within 1–2 months. Standard safety precautions should be used when preparing reagents and handling the polymer for capillary electrophoresis. Care should be taken to avoid any human or other biological contamination to the samples during processing. The use of laminar flow hoods commonly used for PCR techniques is recommended when performing the enzymatic and PCR steps for AFLP analysis.

AFLP analysis was performed with an AFLP Plant Mapping Kit (Applied Biosystems, Inc.) (19). This kit includes the following reagents: pre-selective PCR primers, selective PCR primers, core

mix for PCR amplification, positive PCR amplification control DNA, and DNA adaptors. The following reagents are not included and must be purchased separately: enzymes MseI (concentration 50 units/µL; New England Biolabs Inc. (NEB)), EcoRI (concentration 50 units/µL; Gibco BRL) and T4 DNA ligase and ligase buffer with ATP (concentration 5 Weiss units/µL; Gibco BRL), 10× Running Buffer (Applied Biosystems, Inc.), internal ROX size standard (Applied Biosystems, Inc.), POP7 polymer (Applied Biosystems, Inc.), Hi-Di formamide (Applied Biosystems, Inc.), bovine serum albumin (BSA, concentration 5 mg/mL (Gibco BRL)), Tris-EDTA (TE) buffer (concentration 20 mM Tris–HCl, 0.1 mM EDTA, pH 8; molecular biology grade reagents from Sigma), and sodium chloride (NaCl; concentration 0.5M; molecular biology grade reagent from Sigma). Modifications (20) to the manufacturer's recommended procedure (19) were the following:

1. To reduce pipetting error with small volumes of reagents, master mixes were prepared for enzymes and adaptors for the digestion–ligation step in the kit. Master mixes were created, vortexed, and used at the recommended volumes for no less than 20 samples even if there were only two actual samples to process. This step increases consistency of DNA profiles substantially. All steps were carried out with enzymes maintained in a cold block (–20°C) or on ice while reagents were being prepared.

2. Speed and accuracy in preparation of the reagents greatly improved the consistency of the AFLP result. It is our recommendation that duplicate sample processing be performed by the same or different analysts for confidence in the results while training with the AFLP method.

3. Positive (known DNA from any species for use as a batch standard) and negative (reagents and water but no DNA) controls should be used with each batch of samples processed and analyzed to increase confidence in the results. The kit comes with a positive corn DNA control for the PCR amplification step (see Note 4). For the previous digestion–ligation steps, it is recommended that a laboratory standard be generated for your laboratory and used consistently in addition to the corn control. The negative control sample should be volume adjusted with sterile water so that all reaction volumes are equal. In addition, the personnel that are processing the AFLP samples should be genotyped by AFLP as elimination standards for contamination as the AFLP method is not species-specific.

4. All buffers and polymers should be taken from the freezer or refrigerator and warmed to room temperature (approximately 30 min) prior to use on the 3130 DNA sequencer. This allows for increased consistency in results.

5. For genetic analysis, a custom bin set must be generated with the Genemapper software for automated comparisons of DNA

samples in a database. This is best done with the technical assistance of the software supplier (Applied Biosystems, Inc.; Foster City, CA). It does not matter which bin sets you choose to score the presence or absence of DNA fragments, however, to compare or combine AFLP database information, you must be using common bin sets to share data between laboratories.

3. Methods

1. *Sample preparation.* DNA samples of high molecular weight and the laboratory positive control DNA should be diluted or concentrated to approximately 20 ng in a final volume of 5 μL of water in a thin-walled 0.2 mL PCR tube.

2. *Digestion–ligation enzymatic reaction.* The adaptor pairs provided in the kit should be heated to 95°C for 5 min, and then cooled to room temperature for 10 min to prevent adaptor-to-adaptor annealing which could result in tandem adaptors being accidently amplified as an inadvertent template source. Centrifuge for 10 s at $14,000 \times g$.

3. The enzyme master mix was prepared by combining 20× (1 unit MseI, 5 units EcoRI, and 1 Weiss unit ligase) in a 0.5 mL sterile tube and placing in a −20°F cold block.

4. The adaptor master mix was prepared by combining 20× (2.2 μL 5× ligase buffer, 1.1 μL 0.5M NaCl, 0.11 μL BSA, 1 μL Mse1 adaptor and 1 μL EcoRI adaptor). The volume was then brought to a final volume equal to 20× (6 μL minus the volume of the enzyme mix) with distilled water.

5. The entire adaptor mix was transferred to the tube containing the enzyme master mix, vortexed gently to mix. Centrifuge for 10 s at $14,000 \times g$.

6. 6 μL of the enzyme/adaptor mix was added to each tube containing DNA and gently mixed. The 0.2 mL tubes were incubated at 37°C in a thermal cycler with a heated lid for 2 h.

7. *Pre-selective PCR amplification reaction.* Dilute each sample from step 6 above with 189 μL of TE buffer. Gently mix and transfer a 4 μL aliquot to a new 0.2 μL PCR tube and set aside. The negative control can be set to an equivalent volume with 4 μL of TE buffer with no DNA.

8. Prepare a master mix by combining [the total number of samples + 2 additional samples] × [1 μL AFLP EcoRI and MseI pre-amplification primers + 15 μL AFLP core mix]. Transfer 16 μL of the pre-selective master mix to the 4 μL diluted sample in 0.2 μL PCR tube and gently mix.

9. PCR amplify these samples with the following settings:
   (a) 72°C, 2 min hold step.
   (b) 20 cycles×[94°C/20 s, 56°C/30 s, 72°C/2 min].
   (c) 60°C, 30 min hold step.
   (d) 4°C storage.

10. Transfer these samples to 0.5 mL sterile tubes and dilute again with 380 μL TE buffer.

11. *Selective PCR amplification.* If you are beginning this process with a new plant species, then the kit provides 64 possible selective PCR primer combinations to test against your samples to select the optimal primer sets to generate sufficient numbers of well-spaced and easily scorable DNA fragments with a size range of 50–500 nucleotide bases. For marijuana, we used the selective primer pairs in the following combinations: (1) EcoRI-ACT FAM + MseI-CAA, (2) EcoRI-ACT FAM + MseI-CAT, (3) EcoRI-AAG JOE + MseI-CAT, and (4) EcoRI-AAG JOE + MseI-CTA. These primer sets gave fragments that met our scoring criteria for the initial 100 marijuana samples tested that yielded consistent duplicate DNA profiles.

12. 3 μL of the diluted reaction from step 10 above was transferred to a 0.2 mL tube and set aside.

13. A master mix for selective amplification was prepared using any of the primer sets individually listed in step 11. The master mix was prepared using the formula [total number of samples + two additional samples]×[1 μL of the AFLP EcoRI primer, 1 μL of the AFLP MseI primer, 15 μL core mix].

14. Transfer 17 μL of the master mix to each of the tubes containing 3 μL of sample and gently mix.

15. Samples are then PCR amplified using the following "touch-down" PCR method:
   (a) 94°C/2 min hold step.
   (b) 1 cycle×[94°C/20 s, 66°C/30 s, 72°C/2 min]
   (c) 1 cycle×[94°C/20 s, 65°C/30 s, 72°C/2 min]
   (d) 1 cycle×[94°C/20 s, 64°C/30 s, 72°C/2 min]
   (e) 1 cycle×[94°C/20 s, 63°C/30 s, 72°C/2 min]
   (f) 1 cycle×[94°C/20 s, 62°C/30 s, 72°C/2 min]
   (g) 1 cycle×[94°C/20 s, 61°C/30 s, 72°C/2 min]
   (h) 1 cycle×[94°C/20 s, 60°C/30 s, 72°C/2 min]
   (i) 1 cycle×[94°C/20 s, 59°C/30 s, 72°C/2 min]
   (j) 1 cycle×[94°C/20 s, 58°C/30 s, 72°C/2 min]
   (k) 1 cycle×[94°C/20 s, 57°C/30 s, 72°C/2 min]
   (l) 20 cycles×[94°C/20 s, 56°C/30 s, 72°C/2 min]

      (m) 60°C/30 min hold step.

      (n) 4°C, storage.

16. *Preparation of samples for detection*. Final PCR product (1 μL) is added to 23 μL Hi-Di formamide + 1 μL of ROX size standard (50–500 size range) to prepare for loading the sample into the 3130 DNA sequencer (see Note 5). Each sample is heated to 95°C for 3 min, quick chilled on ice, and placed in the sample tray for injection. Any remaining sample can be stored at –20°C until the DNA profile has been collected and verified as acceptable.

17. Electrophoresis conditions for size separation of DNA fragments are the following:

    (a) For simple DNA patterns: 5 s injection time, 13 kV injection voltage, 26 min run time.

    (b) For more complex DNA patterns: 12 s injection time, 15 kV injection voltage, 30 min run time.

18. *Data management*. DNA profiles generated by AFLP are represented as fluorescently labeled peaks that size between 50 and 500 nucleotide bases. The presence or absence of variable peaks in user-designated bin sets in Genemapper 4.0 was originally determined by analysis of 100 different marijuana AFLP profiles. Bins of varying size were created around common DNA fragments and each fragment was sized by comparison to the internal ROX size standard as no allelic ladder is commercially available for sizing AFLP fragments. A score of "1" was given to a peak that fell within the bin and a score of "0" was given if no peak was present in the bin of interest. By scoring DNA profiles in this fashion, a binary code representation of a complex DNA profile can be entered into a database and automated for candidate matching. The bin set and scoring criteria is necessarily user defined as there will always be the possibility that any given sample will contain a peak that falls between the designated bins. As long as the data scorer uses consistent criteria, this is not a significant issue for data management (see Note 6).

4. Notes

1. Collection of plant evidence, especially drug seizures such as marijuana should be collected, documented, and preserved using all of the established standard forensic science procedures. This method has been validated for forensic use (21). Care should

be taken to avoid human contamination of the plant sample; this is essential as the AFLP method is not species-specific. While this feature makes the AFLP method highly valuable and universal to all species; it can be easy to cross-contaminate the sample without proper precautions (20).

2. To decontaminate a sample, a freshly prepared 10% bleach solution (sodium hypochlorite) should be used to soak the seed or plant sample for 2–3 min with gentle agitation. This should be quickly followed by several sterile water rinses to remove any residual bleach solution. The surface of the sample should now be decontaminated (20).

3. AFLP analysis is highly successful on fresh plant samples and seed samples that yield high quality, high molecular weight DNA using any of a variety of plant DNA extraction kits currently commercially available (20).

4. Each of the AFLP steps can be tested to be sure that they are functioning properly by running proper controls during the analysis steps. If an enzyme is not functioning, then the internal positive control set up by the laboratory for the digestion–ligation step should be inconsistent from sample batch to sample batch; otherwise, the same DNA profile should be generated every time. To test the PCR kit reagents for functionality in the pre-selective and selective PCR steps, the corn control supplied with the kit is sufficient and should generate the same DNA profile every time. Negative controls at each step of the process should exhibit only background fluorescence or free fluorescent primers but no DNA peaks should be visualized in the 50–500 nucleotide base size range (19, 20).

5. The quality of the formamide is critical in the detection step. We recommend using the Hi-Di formamide since it is supplied as a deionized quality controlled product. It should be aliquoted in 500 μL aliquots and stored frozen prior to use. Poor quality formamide inhibits proper denaturation of the sample and results in poor resolution of data that can make the sample difficult to score.

6. Candidate matching AFLP samples that are matched against a binary code in a database could also be assessed by viewing the original data files. There is an overlay feature in the Genemapper software that allows for one sample profile to be directly overlaid on the second profile to determine if they are an exact match for number of peaks, both within and outside of the user-designated bins. This is a nice confirmation feature to supplement the binary coding of DNA samples.

Acknowledgments

References

1. Office of National Drug Control Policy (2007) Marijuana on public lands. http://WhiteHouseDrugPolicy.gov. Accessed 07 January 2011

2. Miller Coyle H, Palmbach T, Juliano N, Ladd C, Lee HC (2003) An overview of DNA methods for the identification and individualization of marijuana. Croat Med J 44(3):315–321

3. Howard C, Gilmore S, Robertson J, Peakall R (2009) A *Cannabis sativa* STR genotype database for Australian seizures: forensic applications and limitations. J Forensic Sci 54(3): 556–563

4. Alghanim HJ, Almirall JR (2003) Development of microsatellite markers in *Cannabis sativa* for DNA typing and genetic relatedness analyses. Anal Bioanal Chem 376(8):1225–1233

5. Hurley JM, West JB, Ehleringer JR (2010) Stable isotope models to predict geographic origin and cultivation conditions of marijuana. Sci Justice 50(2):86–93

6. Fry NK, Savelkoul PH, Visca P (2009) Amplified fragment-length polymorphism analysis. Methods Mol Biol 551:89–104

7. Gaggiotti OE (2010) Bayesian statistical treatment of the fluorescence of AFLP bands leads to accurate genetic structure inference. Mol Ecol 19(21):4586–4588

8. Howard C, Gilmore S, Robertson J, Peakall R (2008) Developmental validation of a *Cannabis sativa* STR multiplex system for forensic analysis. J Forensic Sci 53(5):1061–1067

9. Hsieh HM, Hou RJ, Tsai LC, Wei CS, Liu SW, Huang LH, Kuo YC, Linacre A, Lee JC (2003) A highly polymorphic STR locus in *Cannabis sativa*. Forensic Sci Int 131(1):53–58

10. Mendoza MA, Mills DK, Lata H, Chandra S, ElSohly MA, Almirall JR (2009) Genetic individualization of *Cannabis sativa* by a short tandem repeat multiplex system. Anal Bioanal Chem 393(2):719–726

11. Gilmore S, Peakall R, Robertson J (2003) Short tandem repeat (STR) DNA markers are hypervariable and informative in *Cannabis sativa*: implications for forensic investigations. Forensic Sci Int 131(1):65–67

12. Ergül A, Kazan K, Aras S, Cevik V, Celik H, Söylemezoğlu G (2006) AFLP analysis of genetic variation within the two economically important Anatolian grapevine (*Vitis vinifera* L.) varietal groups. Genome 49(5):467–475

13. Jakse J, Kindlhofer K, Javornik B (2001) Assessment of genetic variation and differentiation of hop genotypes by microsatellite and AFLP markers. Genome 44(5):773–782

14. Kuhlman LC, Burson BL, Stelly DM, Klein PE, Klein RR, Price HJ, Rooney WL (2010) Early-generation germplasm introgression from *Sorghum macrospermum* into sorghum (*S. bicolor*). Genome 53(6):419–429

15. Opara UL, Jacobson D, Al-Saady NA (2010) Analysis of genetic diversity in banana cultivars (Musa cvs.) from the South of Oman using AFLP markers and classification by phylogenetic, hierarchical clustering and principal component analyses. J Zhejiang Univ Sci B 11(5):332–341

16. Rahman MH, Rajora OP (2002) Microsatellite DNA fingerprinting, differentiation, and genetic relationships of clones, cultivars, and varieties of six poplar species from three sections of the genus *Populus*. Genome 45(6):1083–1094

17. Rajora P, Rahman H (2003) Microsatellite DNA and RAPD fingerprinting, identification and genetic relationships of hybrid poplar (*Populus* x *canadensis*) cultivars. Theor Appl Genet 106(3):470–477

18. Stefanowicz-Hajduk J, Filipowicz N, Kosinski I, Ochocka JR (2009) Capillary electrophoresis with dual laser detection in separation of amplified fragment length polymorphism fragments. J Sep Sci 32(20):3539–3543

19. (2010) AFLP Plant Mapping Protocol. Applied Biosystems, Inc. Foster City, CA

20. Miller Coyle H (2005) Forensic botany: principles and applications to criminal casework. CRC Press, Boca Raton, FL

21. Miller Coyle H, Shutler G, Tully L, Pagliaro E, Harper A, Palmbach T, Lee HC (2005) Validation of a DNA method for the individualization of plant evidence. http://www.ncjrs.gov/pdffiles1/nij/grants/211999pdf. Accessed 08 January 2011

Chapter 18

Capillary Electrophoresis of Multigene Barcoding Chloroplast Markers for Species Identification of Botanical Trace Evidence

Gianmarco Ferri, Beatrice Corradini, and Milena Alù

Abstract

The analysis of nonhuman biological evidence both animal and botanical to find out the correct species of a sample comes as a great help to crime investigators. Particularly, forensic botany may be useful in many criminal and civil cases, e.g., for linking an individual to a crime scene or physical evidence to a geographic location, or tracking marijuana distribution patterns.

Despite many molecular techniques for species identification so far applied, botanical evidences are still overlooked by forensic scientists due to the lack of reproducible and efficient protocols standardized across a wide range of different organisms and among different laboratories.

Recently, the term "DNA barcoding" has been coined to describe the use of a short gene sequence from a standardized region of the genome as a molecular tool for species identification. DNA barcodes have been successfully applied to a number of animal groups and introduced in forensic science with the application of the mitochondrial gene COI. Building on this success, ongoing investigations have searched for the best barcode to apply to all land plants. Here we describe the basic protocol based on amplification and sequence analysis of barcoding markers for land plants considering the latest developments of Plant DNA barcoding Project. The aim of this chapter is to provide forensic scientists an accurate and reliable tool for assigning unidentified botanical specimens to the correct species as powerful mainstay in investigations, increasing the contributions from nonhuman DNA to forensics.

Key words: Forensic species identification, Forensic botany, DNA barcode, DNA sequencing, Universal primers

1. Introduction

Molecular species identification, involving plant or animal samples, has become a valuable tool in forensic science, both for crime scene investigation and wildlife law enforcement. In genomic era, different methods have been developed (before immunochemistry, biochemical test, and tissue and cells investigation) based on sequence analysis

(1–3), but the use of molecular markers was recently standardized on a large scale by an international taxonomic consortium (Consortium for the Barcode of Life, CBOL) through the use of one short (relatively) DNA region called "BARCODE" (4). A DNA sequence of 648 bp (base pair) of the mitochondrial gene cytochrome c oxidase sub I (COI) has been accepted for animal kingdom because the fixed criteria of universality, coverage, sequence quality, and discrimination of species were satisfied (5).

In agreement with forensic quality procedure and standardization, the Barcode identification system was applied in forensic animal species identification (6, 7) and a new Barcode section was opened on GenBank with specific submission tool (8), because DNA banking is a fundamental prerequisite for identification.

The Barcode approach also has great potential for identifying plants, but faces different challenges when applied to this group (we start remembering that the number of species was globally different, ~420,000 species of plants vs. ~50,000 species for animal vertebrates) (9–11).

The ubiquitous presence of plants makes botanical trace evidence useful for many aspects of criminal investigation (12), but remains underused in forensic due to a lack of experience and botanical knowledge among investigators, evidence-collection teams, and prosecutors and to the difficulty in routinely identifying trace material by traditional morphological methods. The information generating from plant DNA identification can be used to provide links between crime scene and individuals, test alibis, verify a sample's geographic origin, ascertain the possession of illegal or endangered species, and much more (13–15).

Differently from animal, the attempts to identify a single barcoding locus for plants have largely been unsuccessful, as the slow evolutionary rate of plant mitochondrial DNA did not allow the use of COI marker and now it is generally agreed that only a multilocus approach based on plastid (equivalent to mtDNA source for plant genome) data is the most effective strategy. A variety of loci have been recently suggested in both the nuclear and plastid genome (16–18), and in 2009, the Barcode Consortium approved a two core-locus combination (rbcL+matK, the trnH-psbA marker should be collected as a back-up to matK) for land plants identification (19).

Before the barcode executive committee final evaluation, we selected two noncoding plastid regions (trnL-F and trnH-psbA) among the 7–10 candidate markers to develop a molecular-based identification system for forensic purpose which provides criteria to progressively identify an unknown plant sample to a given taxonomic rank (20). We analyzed 63 individual plants which usually grow in our region and, considering all the species examined, we were able to identify at species level more than 60% of samples. The major drawback encountered was related to the poor coverage of species in

sequence database and the lack of authenticate reference DNA sequence. Consequently, given the development of barcoding initiative and the ability of this International Consortium of processing thousand of reference sequence in a short time, we review our previous developed protocol by analyzing the same collected samples with the approved barcoding markers (rbcL and matK) by comparing the different use and the different performance when applied in the forensic field (manuscript in preparation). We remember that forensic botany is underused and most potential users of this method are not botanical experts, so that the characteristics of reproducibility and standardization are fundamental to accurately identify the specimens. The final step consists in matching the sequence of the evidence sample to a reference sequence through query of public nucleotide repositories. Basic Local Alignment Search Tool (BLAST) web server at NCBI provides the easiest way to run a quick search through online query submission against the largest existing and constantly updated database of sequences (21, 22).

Since the use and the application in a forensic contest are not immediately intuitive, we take advantage of this protocol to describe accurately every step to perform the database search.

During the commission of outdoor crimes, plant material may be frequently transferred from the crime scene to the victim or perpetrator; even if identification to the species level is not always possible, the identification to higher taxonomic rank can be helpful in many situations where the application of a broad species concept is accepted.

The developed protocol was the same independently from used markers (we discuss the best markers' uses in the next paper in preparation) and the approach described here shows all the four-pair of primers, but the user should consider that the discrimination success does not increase after the use of two or three loci.

2. Materials

2.1. DNA Extraction

1. QIAamp DNeasy Plant Mini-kit (Qiagen, Hilden, Germany).
2. 1M DTT Dithiothreitol (SIGMA-Aldrich), in 10 mM Sodium Acetate, pH 5.2.
3. 1.5 mL Safe-Lock tubes (Eppendorf AG, Germany).

2.2. PCR Amplification

1. All primers were selected from the literature (16, 19, 20). PCR/sequencing primers for *TrnL-TrnF* intergenic spacer (amplicon between bp 49868–50305 of *Nicotiana tabacum* NC_001879 chloroplast genome, including primers).

   TrnL_F 5′-GGTTCAAGTCCCTCTATCCC-3′, 0.2 μM

   TrnF_R 5′-ATTTGAACTGGTGACACGAG-3′, 0.2 μM

PCR/sequencing primers for *PsbA-TrnH* intergenic spacer (amplicon between bp 28–589 of *N. tabacum* NC_001879 chloroplast genome, including primers).

PsbA_F 5′-GTTATGCATGAACGTAATGCTC-3′, 0.2 μM.

TrnH_R 5′-CGCGCATGGTGGATTCACAATCC-3′, 0.2 μM.

PCR/sequencing primers for *matK gene* (amplicon between bp 173–1046 in Arabidopsis thaliana gene sequence, including primers):

MatK_3F 5′-CGTACAGTACTTTTGTGTTTACGAG-3′, 0.2 μM.

MatK_1R 5′-ACCCAGTCCATCTGGAAATCTTGGTTC-3′, 0.2 μM.

PCR/sequencing primers for *rbcL gene* (amplicon between bp 1–599 in Arabidopsis thaliana gene sequence, including primers):

RbcL_F ATGTCACCACAAACAGAGACTAAAGC.

RbcL_R GTAAAATCAAGTCCACCRCG.

All primers were purchased from OPERON-MWG biotech, desalted and lyophilized.

2. QIAGEN Multiplex PCR kit (Qiagen).

3. MicroAmp reaction tubes (0,2 μL) (Applied Biosystems).

4. Thermal Cycler (Veriti thermal cycler or GeneAmp 9700 PCR system) (Applied Biosystems).

5. BSA fraction V (0.04% in PCR) (Sigma Aldrich).

2.3. PCR Purification

1. Exonuclease I and Shrimp-Alkaline Phosphatase, ExoSAP-IT (GE healthcare, Bucks, UK).

2.4. Electrophoresis

1. 2% Agarose (Sigma Aldrich).

2. TBE: Tris-base, Boric Acid, EDTA pH 8.00 (Sigma Aldrich).

3. Ethidium bromide (0.5 μg/mL) (Sigma Aldrich).

4. UV light transilluminator (254–366 nm).

2.5. Cycle Sequencing and Postsequencing Purification

1. BigDye Terminator Cycle Sequencing Kit v 1.1 (Applied Biosystems).

2. Sequencing primers, F and R (see above).

3. Thermal Cycler (2720 AB).

4. DyeEx 2.0 Spin Kit (QIAGEN).

5. Eppendorf Microcentrifuge MiniSpin Plus.

2.6. Capillary Electrophoresis

1. ABI PRISM 3130 DNA Genetic Analyzer (Applied Biosystems).

2. HiDi formamide (Applied Biosystems).

3. Performance Optimizer Polymer 4 (Applied Biosystems).

4. 36 cm array (Applied Biosystems).

5. Applied Biosystems DNA Sequencing Analysis Software V5.2.

3. Methods

3.1. DNA Extraction

Perform DNA extraction starting from ~1.0 g of tissue (fresh or dried) with the QIAGEN DNeasy Plant mini-kit. Disrupt the tissue directly in lysis buffer and add 10 µL of 1M DTT solution, then follow the manufacturer's guidelines (see Notes 1 and 2).

3.2. Amplification

1. Set up PCR for amplification of target sequence by mixing the following reagents in a total reaction volume of 10 µL using Qiagen Multiplex PCR Kit:

Reagents	Volume per sample
2× Qiagen PCR Master Mix	5 µL (see Note 3)
10× Primer Mix	1 µL (see Notes 4 and 5)
BSA	0.2 µL
ddH$_2$O sterile	1.3 µL
Template DNA (1–5 ng/µL)	2.5 µL

2. Amplify using the following thermal cycling conditions: an initial cycle at 95°C for 15 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 55°C 1 min, extension at 72°C for 45 s, and a final extension step at 72°C for 10 min (see Notes 6 and 7).

3.3. Agarose Gel Analysis

Run the PCR products at approximately 100–150 V in 2% agarose gel (TBE buffer, Tris-borate-EDTA) containing 0.5 µg/mL of ethidium bromide with a molecular weight markers to confirm the amplification and to enable analysis of correct single fragment size. Examine the bands corresponding to amplified fragments using a UV transilluminator (254–366 nm) and record by taking a Polaroid photograph.

The size of PCR products according to 63 species previously analyzed ((20) manuscript in preparation) could be highly variable (see Note 8):

psbA-trnH = from 268 to 740 bp

trnL-trnF = from 173 to 608 bp

rbcL = from 553 to 554 bp

matK = from 825 to 843 bp

3.4. PCR Purification

Remove the remaining primers and unincorporated dNTPs by incubation of 5 µL of PCR product with 2.5 µL of ExoSAP enzymes for 15 min at 37°C followed by inactivation by heating at 80°C for 15 min.

3.5. DNA Sequencing

Single PCR products can be sequenced using both the forward and reverse PCR primer (the same used for PCR reaction).

Perform sequencing reaction on a dedicated thermal cycler (AB 2720) in a total volume of 20 µL with 2–4 µL of cleaned PCR product, 4 µL of BigDye terminators v.1.1, 2 µL of 10× Sequencing buffer, 2 µL of 10× primer solution (F or R).

Carry the cycle sequencing reaction using the following conditions: initial denaturation at 96°C for 1 min following 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 a 4°C.

The samples are now ready for purification.

3.6. Postextension Treatment

The DNA sequencing product was treated with the DyeEx 2.0 Spin Kit to remove the unincorporated primers following manufacturer's guidelines by centrifuging the spin column at 3,500 rpm in a Eppendorf MiniSpin plus (see Note 9).

3.7. Capillary Electrophoresis and Analysis of Sequencing Products

Purified products were resolved by electrophoretically separation in a ABI 3130 DNA sequencer. For preparation of samples to electrophoresis separation, 1 µL of sequencing purified product was mixed to 12–15 µL of formamide and run on an 3130 genetic analyzer using POP 4. Data were analyzed using AB DNA Sequencing Analysis Software V5.2.

3.8. Database Search

Hereafter are described the main steps to compare the sequence of interest with NCBI library of sequences performing a BLAST search via web interface.

1. Point the browser to access the home page of the National Center for Biological Information (NCBI) BLAST server at http://www.ncbi.nlm.nih.gov/blast.

2. In the "Basic BLAST" section, choose to run *nucleotide BLAST* program which enables to search a nucleotide database using a nucleotide query. The following page is organized in five sections: Enter Query Sequence, Choose Search Set, Program Selection, (submit) BLAST, and Algorithm parameters.

3. At the top of the web page in the "Enter Query Sequence" section, paste the query sequence in the search text area (see Note 10).

4. After entering the query sequence, select the sequence database to run the search with "Choose Search Set." First check the "Others" button and then use the pull-down menu to choose "Nucleotide collection (nr/nt)."

5. In the "Program Selection" menu, choose optimize for "Somewhat Similar Sequences (blastn)" as BLAST algorithm to adopt for search and alignment (see Note 11).

6. Set the search parameters of the selected BLAST algorithm as default options (see Note 12).

7. Click on the BLAST button to submit the job. The query sequence will be compared to all of the entries in the specified database (23).

8. BLAST will now open a new window (*Job running page*) which reports the status of the running job and an estimate of completion time.

9. In few minutes, the computed results will be presented in the *output* window. For details on the content of the output report, see Note 13.

10. The first entry in the output file represents the nucleotide sequence available in the database which displays the best similarity with the query sequence (best close match). The objective way to assess the significance and reliability of the reference match (e.g., whether the alignment portrays possible biological relationship or it is attributable to chance) is through the statistical values produced for each alignment pair, such as the Scores and Expectation Value (E-value). That is, the higher the score and the lower the E-value, the more significant the hit. However, there are situations in which interpretation of results may be complicated by several factors and following these rules may not lead to a biologically meaningful correspondence (24). The different possible outcomes that may be encountered in species identification attempt of unknown botanical evidence in publicly accessible database are illustrated in a simplified way together with some basic indications in Table 1 (see Note 14).

Table 1
Outcomes that may be encountered in species identification searching in sequences databases

Situation	Description	Outcome
The searched sequence doesn't already exist in the database	The genomic region for a given barcode is not yet available for comparison	Ambiguous or incorrect (false-positive) ID as comparison is only achievable at higher taxonomic ranks (genus, family, order)
The searched sequence exists in the database		
Single best match with exact sequence	query sequence is returned as the unique best hit with the highest statistical values	Correct ID at *species* level
Multiple best matches with exact sequence	query sequence plus one or more identical sequences are returned as best hits with highest statistical values	Correct ID only at *genus* or *family* level (based on the taxonomic category giving the best matches)
Best match with different unrelated sequence(s)	the true species is present in the reference alignment, but the method failed to assign as best hit	Incorrect ID

ID identification

11. To facilitate the sequence comparison and the species identification, specific databases are now in preparation (www.barcodinglife.com) (25) (see Note 15).

4. Notes

1. Liquid Nitrogen is not always easily available in a forensic lab. We tried to eliminate this step also by adding DTT. Even if the tissue disruption was not complete, the high copy number of chloroplast DNA leads to a successful analysis even from low and degraded DNA.

2. Purified DNA was eluted from DNeasy spin column using 50 µL of AE buffer.

3. The QIAGEN Multiplex PCR kit contains a master mix with preoptimized concentrations of HotStarTaq DNA Polymerase, $MgCl_2$ dNTPs, and PCR buffer with the factor MP to ensure comparable efficiencies for annealing and extension of all primers without the need of optimization and should be used with standard-quality primers.

4. All primer stock solutions should be normalized to a concentration of 100 µM using sterile ddH_2O or TE buffer and stored at –20°C.

5. The 10× Primer mix should be prepared in a 250 µL of total volume by adding 5 µL of 100 µM stock solution of each primer (the 10× primer mix contain each primer at 2 µM).

6. It was confirmed with *matK* locus the lack of availability of robust primer sets for all land plants and especially for gymnosperms (19).

7. PCR standard conditions apply widely with high amplification success to rbcL, trnH-psbA, trnL-trnF barcodes, whereas matK requires modified protocols to allow successful amplification in a substantial proportion of samples.

8. Negative amplification and product showing more than one band or aspecific bands in agarose gel were reamplified, respectively, by adding more DNA and by more stringent PCR condition.

9. Some artifact peaks can appear in the electropherogram if the spin protocol does not remove efficiently the unincorporated dye terminators. In this case, the purification can be repeated on the same product.

10. The BLAST web page accepts input sequences in a number of different formats. Accepted input types are IUPAC FASTA or bare sequence format (lines of sequence data, without the

FASTA definition line) or sequence identifiers such as NCBI Accession Number, GenBank Identification Number (GI). Instead of manual sequence entry into the text box, it is possible to use the browse button to upload a file from your local disk. The file may contain a single sequence or a list of sequences. The data may be either a list of database accession numbers, NCBI gi numbers, or sequences in FASTA format. Furthermore, a descriptive title (Job Title) may be entered to obtain a unique identifier for each search. The optional "Query Subrange boxes" limits the search to a subrange of the query sequence.

11. In the nucleotide search, the alignment comparison between query sequences and database entries may be provided by alternative algorithms, MegaBLAST, discontiguous MegaBLAST, and the standard Blastn. They differed in the default parameters governing the sensitivity of searches. Generally, Blastn provides more sensitive and less stringent alignments at a price of slower computed speed, whereas MegaBLAST is specifically designed for more selective and faster alignments with long sequences. For a more exhaustive research, it is recommended to perform the analysis with both algorithms and then compare the results returned.

12. Using the default settings for a BLAST search, regardless of the algorithm implemented, is a sensible approach because they are set up to give the best all-round results and in most cases they don't need to be changed. Moving beyond the default conditions may depend on the type of a specific strategy and it may be necessary to improve sensitivity and selectivity in some specific applications. To change search parameters, it is recommended to consult NCBI Educational Resources at http://www.ncbi.nlm.nih.gov/Education/.

13. BLAST provide a three-section output report: on the top a header with query sequence information and details on the work submitted (*Job identifier*) followed by a summarized graphical overview with the hits found, a hit table showing matching sequence descriptions, and finally the pairwise alignments for the sequence of interest with database records. In the graphic summary, the distribution of Blast Hits on the Query Sequence is displayed as colored lines according to the similarity score from red (the highest) to black (the lowest). Mousing over an alignment shows the alignment definition and score in the box at the top and clicking an alignment displays the alignment detail. In the next section called "Descriptions," the highest scoring alignments ("high-scoring pairs," or HSPs) are present along with several one-line descriptors. The description lines are sorted by default on the basis of E-value, so that the most significant alignments (lowest E values) are at the top. Other indicators provided are GenBank identifier and a short

description of the matching sequence, Max and Total Scores, Query Coverage, and Max Ident. The guiding criterion with which to sort the results could be changed clicking on one of the other indicators. At the bottom of the report page the *Alignment* section, each alignment pair query to hit (which is labeled as *Sbjct*) together with information about the specific alignment are provided. For details see ref. (4).

14. Several contributing factors affect the probability of correct assignment of a sequence and hinder the interpretation phase of BLAST output page when searching an unknown plant sample against a publicly available nucleotide database. First of all, the complexity of plant kingdom (interspecific vs. intraspecific variation) and the large number of plant species compared to animal counterpart (>400,000 species known), the status of the nucleotide database with rate of marker redundancy, and the quality and uniformity of records. Furthermore, also the type of algorithm implemented could influence the output results, as for example BLAST due to its intrinsic properties sometimes rewards higher query coverage rather than higher nucleotide identity with query sequence. These issues warrant careful interpretation of each specific result so that only some simplified situations are explained in Table 1. Anyhow, integration of search over additional databases such as that developed by CBOL (briefly described below) and implementations of a multilocus barcode approach with several markers may improve the chance of reliability of results and aid the decision process to be used when making species identifications.

15. The Barcode of Life Data System (BOLD) is the online workbench that aids collection, management, analysis, and use of DNA barcodes data (5). The BOLD-Identification System (IDS) is the dedicated species diagnosis tool and only recently accepts queries also from the plant kingdom from the candidate barcode genes *rbcL* and *matK*. At the moment, however, there are very few records in BOLD so most queries will likely not return a successful match. The aim of CBOL to support the development of a carefully validated library of reference DNA barcodes sequences from voucher specimens will lead to reliable identification process and to overcome some of the showed troubles.

References

1. Tsai LC, Yu YC, Hsieh HM, Wang JC, Linacre A, Lee JC (2006) Species identification using sequences of the trnL intron and the trnL-trnF IGS of chloroplast genome among popular plants in Taiwan. *Forensic Sci Int* 164, 193–200

2. Tsai LC, Yu YC, Hsieh HM, Liu KL, Linacre A, Lee JC (2008) Bidens identification using the noncoding regions of chloroplast genome and nuclear ribosomal DNA. *Forensic Sci Int Genet* 2, 35–40

3. Ward J, Gilmore SR, Robertson J, Peakall R, (2009) A Grass Molecular Identification System for Forensic Botany: A Critical Evaluation of the Strengths and Limitations. *J Forensic Sci* 54, 1254–1260

4. Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003) Biological identifications through DNA barcodes. *Proc R Soc Lond B* 270, 313–321

5. Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Phil Trans R Soc Lond B* 360, 1805–1811

6. 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

7. Ferri G, Alù M, Corradini B, Licata M, Beduschi G (2009) Species identification through DNA "barcodes". *Genet Test Mol Biomarkers* 13, 421–426

8. Benson D A, Karsch-Mizrachi I, Lipman D J, Ostell J, Sayers EW (2010) GenBank. *Nucleic Acids Res*, 38, D46–D51

9. Cowan RS, Chase MW, Kress JW, Savolainen V (2006) 300 000 species to identify: problems, progress and prospects in DNA barcoding of land plants. *Taxon* 55, 611–616

10. Erickson DL, Spouge J, Resch A, Weigt LA, Kress WJ (2008) DNA barcoding in land plants: developing standards to quantify and maximize success. *Taxon* 57, 1304–1316

11. Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW et al (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Mol Ecol Notes* 9, 130–139

12. Coyle HM, Lee CL, Lee HC, Lin WY, Palmbach TM (2005) Forensic **b**otany: using plant evidence to aid in forensic death investigation. *Croat Med J* 46 (4), 606–612

13. Coyle HM, Ladd C, Palmbach T, Lee HC (2001) The green revolution: botanical contributions to forensic and drug enforcement. *Croat Med J* 42, 340–345

14. Virtanen V, Korpelainen H, Kostamo K (2007) Forensic botany: usability of bryophyte material in forensic studies. Forensic Sci Int 172, 161–163

15. Bock JH, Norris DO (1997) Forensic botany: an under-utilized resource. *J Forensic Sci* 42, 364–367

16. Kress JW, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci USA* 102, 8369–8374

17. Chase MW, Cowan RS, Hollingsworth PM et al (2007) A proposal for a standardised protocol to barcode all land plants. *Taxon* 56, 295–299

18. Fazekas AJ, Burgess KS, Kesanakurti PR et al (2008) Multiple multilocus DNA barcodes from the Plastid Genome discriminate plant species equally well. *PloS ONE* 3, e2802 doi:101371/journalpone0002802

19. CBOL Plant Working group (2009) A DNA barcode for land plants. *Proc Natl Acad Sci USA 106,* 12794–12797

20. Ferri G, Alù M, Corradini B, Beduschi G (2009) Forensic botany: species identification of botanical trace evidence using a multigene barcoding approach. *Int J Legal Med* 123, 395–401

21. Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990) Basic Local Alignment Search Tool. *J Mol Biol* 215, 403–410

22. McGinnis S, and Madden TL (2004) BLAST: at a core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* 32, W20–W5

23. Ladunga I (2009) Finding similar nucleotide sequences using network BLAST searches. *Curr Protoc Bioinformatics* Chapter 3:Unit 33

24. Ross HA, Murugan S, and Li WL (2008) Testing the reliability of genetic methods of species identification via simulation. *Syst Biol* 57(2), 216–230

25. Ratnasingham S, Hebert PD (2007) BOLD: the barcode of life data system. (http://www-barcodinglifeorg) *Mol Ecol Notes* 7(3), 355–364

Part II

Capillary Electrophoresis and DNA Sequencing

Capillary Electrophoresis of Big-Dye Terminator Sequencing Reactions for Human mtDNA Control Region Haplotyping in the Identification of Human Remains

Marta Montesino and Lourdes Prieto

Abstract

Cycle sequencing reaction with Big-Dye terminators provides the methodology to analyze mtDNA Control Region amplicons by means of capillary electrophoresis. DNA sequencing with ddNTPs or terminators was developed by (1). The progressive automation of the method by combining the use of fluorescent-dye terminators with cycle sequencing has made it possible to increase the sensibility and efficiency of the method and hence has allowed its introduction into the forensic field. PCR-generated mitochondrial DNA products are the templates for sequencing reactions. Different set of primers can be used to generate amplicons with different sizes according to the quality and quantity of the DNA extract providing sequence data for different ranges inside the Control Region.

Key words: mtDNA control region, Human identification, mtDNA sequencing

1. Introduction

In the past years, mtDNA analysis has been proven to be a powerful tool for the identification of human remains in cases of severely degraded samples or in cases in which, in the absence of direct relatives, comparisons must be made with maternal relatives. Although labor intensive and time consuming, the sequencing of mtDNA control region PCR products is reliable and the most informative technique with this purpose.

The analysis of mtDNA Control Region (CR) involves two main processes: (1) amplification with specific primers spanning the entire CR from nucleotide positions 16024-576 and (2) the sequencing of the amplicons with sequencing primers and Big-Dye

terminator cycle-sequencing kit. The final analysis is performed by capillary electrophoresis, the chosen method to detect the fluorescent fragments generated during the cycle-sequencing reaction.

The growing knowledge about the mitochondrial DNA peculiarities and the availability of the technology has made the use of this strategy in forensic laboratories with identification purposes possible in two main fields: (1) analysis of degraded samples or low nuclear DNA (nDNA) quantity samples (hair shafts) and (2) identification and kinship analysis of human remains with maternal relatives as reference samples (human remains identification, mass disasters, mass graves, etc.).

The entire process involves the following stages:

1. DNA extraction and quantification.
2. Amplification from DNA extracts and quality control (electrophoresis).
3. Purification of PCR products.
4. Sequencing reaction.
5. Clean-up of sequenced products.
6. Electrophoresis.
7. Sequence analysis and quality control.

The selected strategies regarding extraction and amplification methods will depend on the quality and quantity of the sample but we will not be looking into these strategies in any depth here. Nevertheless, a brief description of amplification is necessary since the sequencing primers depend on the amplification strategy.

Since the first publication about protocols for mtDNA sequencing analysis for forensic application in 1995 (2), several strategies have been developed (3–5). In the first publication only the analyses of Hypervariable Regions 1 and 2 (HV1 and HV2) were addressed: HV1 covering positions 16024–16365 and HV2 covering positions 73–340. Nevertheless, the increasing knowledge about features and particular drawbacks of mtDNA (point and length heteroplasmy, the issue of the population-associated common haplotypes, and match interpretation) has led to the analysis of the complete Control Region and today, polymorphisms between 16024–16569 and 1–576 are often analyzed. Furthermore, expanding the analysis to the coding region offers the possibility of increasing discrimination in cases of common haplotypes, or to perform the correct assignment of a sample to its specific haplogroup when, with the HV1 and HV2 data alone, this is not possible (6–8). Obviously, not all the approaches can be applied to the identification of human remains in a forensic context. Quality and quantity of available samples as well as the capabilities and experience of the laboratory will determine the analytical strategy.

Table 1
Amplification primers and their combination to generate amplicons of different sizes

Primer[a]	5′–3′ sequence	Amplicon size
L15997 H16395 H17	CAC CAT TAG CAC CCA AAG CT (20-mer) CAC GGA GGA TGG TGG TCA AG (20-mer) CCC GTG AGT GGT TAA TAG GGT (21-mer)	L15997/H16395 = 437 bp L15997/H17 = 629 bp
L16555 L48 H408	CCC ACA CGT TCC CCT TAA AT (20-mer) CTC ACG GGA GCT CTC CAT GC (20-mer) CTG TTA AAA GTG CAT ACC GCC A (22-mer)	L48/H408 = 401 bp L16555/H619 = 612 bp
H619 L350	GGT GAT GTG AGC CCG TCT AA (20-mer) GCA CTT AAA CAC ATC TCT GCC A (22-mer)	L350/H619 = 310 bp

[a] Named according to their 3′-end position the rCRS

As far as identification of human remains is concerned, and for a laboratory routinely performing mtDNA analysis, a practical and useful strategy would consist of the analysis of the entire Control Region in both reference and questioned samples whenever possible. For nondegraded, good level DNA samples (saliva or blood reference samples, human remains with good level nDNA) (see Note 1), the amplification of the Control Region can be carried out in one or two reactions. In our lab we usually perform two reactions using the following primers (see Table 1):

Primer pair L15997/H17

Primer pair L16555/H619

For slightly degraded samples or those with less DNA concentration, the amplification can be carried out in three or more reactions in order to generate shorter segments (9, 10). The strategy we follow to analyze these types of samples (moderate degradation or less than 40 pg/μL) is as follows (see Table 1):

Primer pair L15997/H16395

Primer pair L48/H408

Primer pair L350/H619

The validation of the two strategies has been performed in our laboratory with different types of samples.

For more degraded samples or samples with minute amounts of nDNA, the amplification of mini-amplicons can be the only possible approach (see Note 2). In Chapter 23, the strategy for high degraded samples is described in depth.

The yield and success of the amplification reaction can be monitored either by agarose gel electrophoresis or by microchip electrophoresis (see Chapter 28).
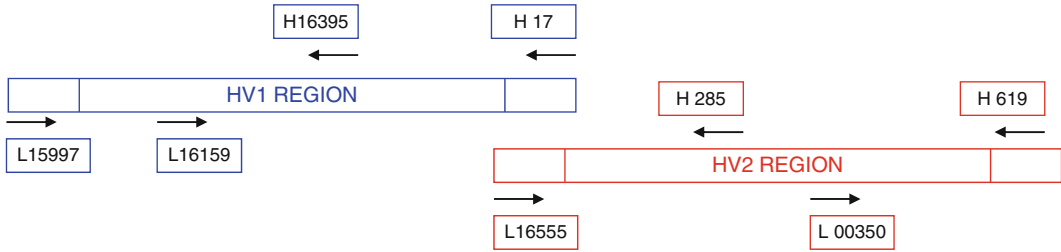
Fig. 1. Sequencing primers used to cover the entire control region generating overlapping fragments (→ forward primer; ← reverse primer).

PCR products must be purified to eliminate primers, dNTPs, and other molecules that could otherwise interfere with sequencing reaction. The purification may be performed with several methods. In our experience, the Qiaquick PCR purification kit (Qiagen) gives good results rendering clean PCR products suitable for subsequent sequencing (see Note 3) (11). During the purification step, a loss of product occurs. DNA yield is about 80% (see Note 4) and this fact must be taken into consideration when estimating the volumes to perform sequencing reactions. Purified PCR products are now ready to undergo sequencing reaction with Big-Dye terminator kit. The reaction Mix contains the primer, Taq polymerase, the four dNTPs, and dye-labeled ddNTPs each with a different dye. During cycle-sequencing, both types of nucleotides are incorporated at random into the growing chain. When a dye-ddNTP is added, the absence of the $3'$-OH precludes the addition of a new dNTP and prevents the elongation. The product of sequencing reactions will contain dye-labeled DNA fragments of each and every size ranging from the primer plus 1 nucleotide to the size of the PCR fragment. Furthermore, as each ddNTP has a different fluorescent-dye, each fragment is dye-labeled with a different color, depending on the incorporated base. Finally, these fragments are separated by electrophoresis and the laser-induced fluorescence is detected. The software installed in the Sequencer allows the data to be collected as electropherograms (EPGs) and performs the analysis of the sequences.

In forensic laboratories, in order to assure the quality of the data, each PCR product is analyzed in forward and reverse strand of the template providing at least two readings for double strand coverage. Furthermore, as the length of fragments extends beyond the maximum clear range of the sequence reading (see Note 5), internal primers are used to generate overlapping sequences to cover the entire region (see Fig. 1).

After the completion of the sequencing reactions, unincorporated dye terminators must be removed from sequencing products otherwise they interfere in the base calling, especially at the beginning of the sequence. Several methods may be used (see Note 6). DyeEx kits in column or 96-well plate format are our routine

methods. Gel-filtration material retains dye terminators, primers, and fragments up to 20 bases while DNA fragments are recovered after a centrifugation step in the flow-through.

After clean-up and subsequent drying, samples are resuspended by using formamide and denatured prior to electrophoresis.

Several platforms are suitable to perform the electrophoresis. We describe here the method for 3130 Genetic Analyzer (Applied Biosystems). This machine works with four capillaries and performs the analysis of four sequencing reactions in 1 h. The data collected during electrophoresis is processed by the analysis software to render an EPG. For each sample (PCR product) SeqScape software performs comparisons and alignments of the sequences generated with each primer to produce a "consensus" sequence. Then, the consensus sequence is automatically compared with the reference sequence (rCRS) (12).

2. Materials

2.1. For Sequencing Reactions

1. Good quality DNA template, in this case purified PCR products.
2. Primers at a concentration of 10 μM (aliquoted and stored at –18°C) (see Note 7).
3. Ultra pure water.
4. Big-Dye terminator ready reaction cycle sequencing kit v1.1 or 3.1 (Applied Biosystems P/N 4337455 and 4337450) (see Note 8).

2.2. For Purification

1. DyeEx 2.0 columns or DyeEx 96 (Qiagen) (see Note 9).
2. Ultra pure water.

2.3. For Electrophoresis

1. Hi DI Formamide (Applied Biosystems, P/N 4311320) (see Note 10).
2. 10× Genetic Analyzer Buffer (Applied Biosystems, P/N 402824) (see Note 11).
3. POP 6 polymer.
4. Capillary array, plates, plate assembly, buffer reservoirs, and accessories provided by the sequencer manufacturer.

3. Methods

3.1. Performing Sequencing Reactions

Sequencing reactions are performed according to the manufacturers' manual instructions (13) with slight modifications. Two versions of the kits are available: v1.1 and 3.1. Both of them work with

Table 2
Primers and quantities of template to perform cycle sequencing

Amplification primers (size)	Sequencing primers[a]	PCR product quantity (total ng)
L15997/H17 (629 bp)	L15997, H17, L16159 and H16395 L16209 and H16164[b]	12–15
L15997/H16395 (437 bp)	L15997 and H16395 L16209 and H16164[b]	5–8
L16555/H619 (612 bp)	L16555, L318 or L350, H619, H285	12–15
L48/H408 (401 bp)	L48, H408 and H285[c]	5–8
L350/H619 (310 bp)	L350, H619	5–8

[a] Nomenclature of primers according to their 3′ position in the rCRS
[b] The use of primers L16209 and H16164 is necessary in case of length heteroplasmy in poly-C stretch surrounding 16189 (see Note 12)
[c] The use of the internal primer H285 is recommended to confirm the readings at the beginning of this segment (see Note 13)

robustness in samples with a minimum amplicon concentration (see below). The kits contain all the required components to perform the sequencing reactions: only the primers are provided by the user. The main points to take into consideration are as follows:

1. *The volume of PCR product.* The quantity of amplified product necessary to perform the sequencing reaction depends on the amplicon size and hence on the regions analyzed. Only samples that meet the estimated required quantity criteria undergo sequencing. The concentration cut-off point, which determines the minimum concentration yielding a good quality sequence, will be established in each laboratory. The product loss must be considered during the purification step. In our experience, the optimal quantity for HV1 and HV2 (437 and 401 bp respectively) is 5–10 ng of PCR product in a final volume reaction of 10 μL. For L15997/H17 and L16555/H619 amplicons, 12–15 ng of PCR products are needed (see Table 2).

2. *The primers.* According to the amplified region, several primers may be used as depicted in Table 2. Sequencing "long" segments involves the use of internal primers in addition to the amplification primers. This strategy allows the reading of the complete segment in both directions. The final concentration in the reaction Mix is 0.4 μM (4 pmol).

Table 3
Typical volumes for cycle sequencing "*n*" samples

Premix	Primer 10 μM	PCR product	Water (qs 10 μL)
4 μL (×*n*)	0.4 μL (×*n*)	2–5.6 μL/sample	0–3.6 μL/sample

3.1.1. Preparing DNA Templates

For a 10 μL sequencing reaction, the sample volume is 5.6–6 μL. DNA templates may need to be diluted if PCR products are too concentrated. Appropriate dilutions need to be made with ultra pure water to obtain the quantities of product specified in the previous paragraph in a volume of 5.6 μL.

3.1.2. Preparing the Mix

A set of primer-specific mixes must be prepared. For a final reaction volume of 10 μL, we need (see Note 14):

1. For each sample and primer: add 0.4 μL of primer to 4 μL of ready reaction premix and mix.
2. Spin briefly and dispense 4.4 μL in each tube.
3. Add 5.6 μL of PCR product previously diluted and mix carefully.
4. Centrifuge tubes to avoid bubbles.
5. Place the tubes in the Thermal cycler (see Note 15).
6. Prepare as many reaction mixes as primers are being used. During the preparation process, keep the premix and primers on ice (see Table 3).

3.2. Purification of Extension Products

1. Place the tubes with extended products at 4°C until the purification step (see Note 16).
2. Follow the manufacturer's protocol (see Note 17) (14).
3. Dry the samples in a vacuum centrifuge without heat (see Note 18).

3.3. Electrophoresis

3.3.1. Preparing Samples for Electrophoresis

1. Dried extension products on tubes (or plate) have to be resuspended thoroughly with formamide (see Note 19). The resuspended products are then transferred to the suitable plates for running on ABI 3130.
2. Add 25 μL of hi-di formamide to each tube.
3. Vortex and centrifuge the tubes (or plate) in order to place the sample at the bottom of the tubes.
4. Denature for 3 min in the thermal block and then place on ice.
5. Transfer the content of each tube to the plate-well according to the order specified on the Sample Manager sheet provided by the running software.

6. Centrifuge the plate to eliminate bubbles and place the sample in the bottom.

7. Place the plate on the plate platform and start electrophoresis.

3.3.2. Electrophoresis Conditions

– 36 cm capillary arrays.

– Running conditions:

Run module: "RapidSeq36_POP6" with "DyeSet/primer DT 3130{BDv3}v1."

Analysis protocol: Basecaller 3130POP6RRv2; Mixed Bases identification >25%; Clear range: manually established depending on the quality of the sequence; Filter Settings: Maximum Mixed bases: 20%; Maximum N's: 10; Minimum clear length (bp): 50; Minimum sample score: 20.

3.4. Analyzing the Data

3.4.1. Inspection of EPGs

Sequencing analysis softwares usually include a tool to automatically check the quality of the electros, but manual inspections are also necessary. A good EPG is characterized by evenly spaced peaks and lack of noise. Peak heights may vary due to different intensities of the fluorophores.

In addition to molecular problems, such as length heteroplasmies, common analytical artifacts can lead to erroneous readings. Some examples are as follows:

– Spikes, produced by urea crystals or bubbles in the polymer. A reinjection of the sample normally solves this problem (see Fig. 2).

– Dye blobs, due to a defective purification.

– Loss of resolution, due to multiple causes such as bubbles in the polymer, defective injections, changes in the running temperature, etc.

3.4.2. Edition of Sequences

Analysis software (Sequencing analysis) performs the analysis of raw data in the form of an EPG. The analysis and edition is performed simultaneously with the ABI PRISM® SeqScape® Software.
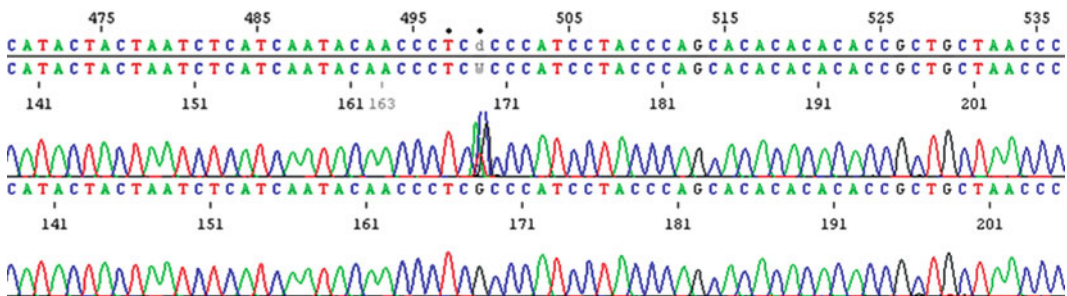
Fig. 2. EPG showing a spike interfering the reading (*top*). The same sample reinjected (*bottom*).

For each sample, the different sequences obtained with direct and reverse primers are aligned and the consensus sequence is compared with the revised mtDNA reference sequence (rCRS). Although the process is automatically done, the visual inspection of the entire length of the EPG must be done to review the data, especially at the beginning and ends of the EPG. Special attention should be paid to identifying point heteroplasmy not detected by the software (see Notes 20 and 21). This revision process is performed by two independent scientists.

Differences from rCRS are shown as a table in the report made by the SeqScape software. To describe the sample's haplotype, nomenclature guidelines and recommendations suggested by the ISFG and EDNAP must be followed (15, 16). This standardization allows different samples to be compared.

3.5. Quality Control and Revision

Unfortunately, errors in mtDNA analysis are more frequent than in nDNA analysis but there is enough literature available describing the types and causes of the errors (17–22). The most obvious problems are mtDNA is more prone to contamination, the mtDNA analysis requires more steps than nDNA, no PCR multiplexes are usually carried out promoting the artificial recombination, and the data is usually transcribed manually to the final reports (allowing for human error). The two main consequences of these errors are (1) false exclusions/no exclusions in the analysis of samples involved in forensic cases and (2) distortion of match probabilities when database searches are performed.

For all the above-mentioned reasons, checking mechanisms in mtDNA data should be applied before considering an mtDNA sequence as definitive. These mechanisms include the following:

1. Checking if all polymorphisms in the haplotypes can be seen in both strands. Samples showing length heteroplasmy must be analyzed using extra internal sequencing primers. Also, polymorphisms located at the beginning or the ends of the amplicons, which are only seen in one strand, must be confirmed using an additional sequencing primer.

2. Reading the sequences twice, by different analysts at two different moments. Only when both readings match the haplotype can it be considered as consensus haplotype.

3. Checking the reading ranges or sequence edition. The reading range is like the name of the genetic marker. If, for some reason, some part of the EPG cannot be unambiguously read (and there is no more sample available for further sequencing), it is mandatory to clearly report the real reading range in that specific sample.

4. Checking if all polymorphisms were previously described. This task is difficult to carry out in some cases but literature and databases are perfect tools to achieve it (23). It is also important to see if the polymorphism was previously described in the haplogroup of the haplotype we are interpreting.

5. Checking point heteroplasmies. In these cases, it is instructive to check the rate of mutation of the nucleotide position where the heteroplasmy is located (24). The mutation rate is not uniform throughout the mtDNA molecule, and there are some positions that are prone to accumulate changes, i.e., hotspots such as 16093 in HVS-I or 152 in HVS-II. A specific mutation rate for each nucleotide position has not been established yet, but some useful information can be gathered in refs. (19, 25, 26). Also, it is important to take into account that the mutation rate varies among different tissues. The possibility of mutation increases in tissues with high level of metabolic activity and depending on the age of the individual (27) (see Note 22).

6. Reading haplotypes from a phylogenetic point of view (assigning the haplotype to a haplogroup). This is useful to avoid some common errors such as artificial recombination or forgotten polymorphisms. The mtDNA phylogeny is continuously improving but the Phylotree web page (http://www.phylotree.org) is a good page to consult as it is continuously updated (28).

7. Checking alternative alignments. Haplotypes can be differently named depending on the alignment with respect to rCRS, mainly in poly-C tracts. This could lead to different results when database searches or comparisons between different labs are performed (29).

8. Performing simple database searches helps to prevent errors. These searches should be carried out even if the statistical evaluation of the results was not necessary (e.g., reference sample not available), that is to say, only for quality control purposes. The EMPOP (http://www.empop.org/) database has proved to be one of the best and more accurate databases publically available (30) and it has a variety of different tools that could assist forensic geneticists in different tasks related to error detection.

9. Finally, if haplotypes cannot be electronically transferred to the final report, they must always be double checked by two independent analysts. Documentation errors are the most frequent type of error and avoiding them only requires being careful (31, 32).

4. Notes

1. Samples with nDNA concentration greater than 50–100 pg/μL are appropriate for this strategy. Lower levels of DNA (5–40 pg/μL) need an increase in the number of cycles from 32 to 36. Nevertheless, in some samples, the amplification is prevented by the presence of inhibitors. In these cases, dilution of the sample may overcome this problem.

2. In this case, the number of amplification cycles is 36. Increasing the number of cycles beyond 36 is not recommended as the risk of false positive and contamination in blanks and negative controls increases.

3. Other filtration methods for PCR clean-up include Microcon 100 (Millipore); Amicon 30 (Millipore); Wizard® SV Gel and PCR Clean-Up System A9281 (Promega Corporation); Illustra MicroSpin S-300 HR Columns (GE); ExcelaPure™ UF PCR Purification System (Edge Bio). An enzymatic digestion with ExoSAP-IT (USB) is also possible.

4. The manual indicates a recovery of 90% but in our experience this value can reach low values as 70–80%.

5. With a 36 cm capillary, well-defined peaks are obtained until 340–350 bases. From this position, uneven peaks and poor resolution sequence are progressively generated.

6. Commercial filtration methods on spin column or 96-well plate: Centri-Sep columns, DyeEx 2.0 or DyeEx 96 (Qiagen), Performa® DTR V3 96-Well Short Plate Kit (EDGE-Bio). Precipitation methods: ethanol-EDTA precipitation or ethanol-EDTA-sodium acetate precipitation. Digestion with SAP and precipitation.

7. Primers 10 μM are aliquoted in small volumes sufficient for 7–8 sequencing reactions. Avoid thawing and freezing cycles. Although not all the primers are equally sensitive to the storage conditions, as a general rule, after 8–10 freeze and thaw cycles aliquots are discarded. During use, primers are kept on ice.

8. Both versions are suitable. V3.1 is better for longer read lengths while v1.1 is better for shorter fragments. Routinely, v3.1 is used in our laboratory. Reaction Mix (400 μL) is aliquoted in 4 tubes × 100 μL to avoid freezing and thawing cycles.

9. As the purification in plate involves the use of multichannel pipettes and the risk of touching the surface of the gel is greater than in columns, DyeEx 96 plate format is only used for reference samples, keeping columns for more critical samples.
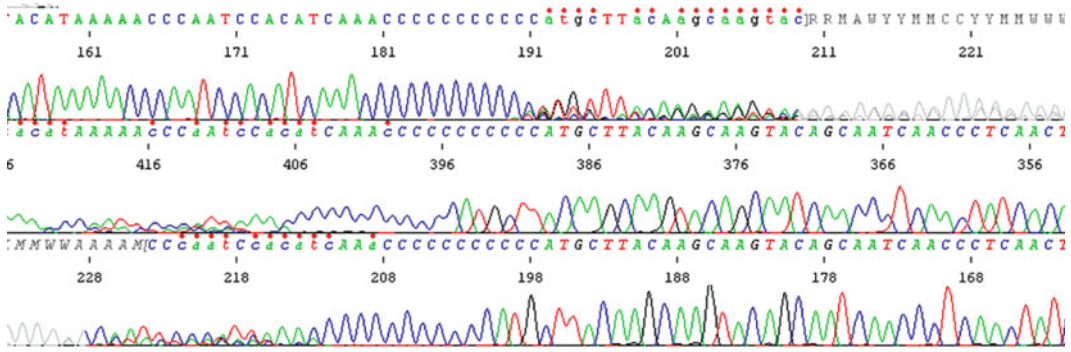
Fig. 3. EPG obtained in a sample with HV1 length heteroplasmy. Primers L15997, H17, and H16395 from *top* to *bottom*.

10. Formamide is highly toxic, mutagenic, and teratogen. Formamide causes eye, skin, and respiratory tract irritation. It is a possible reproductive and birth defect hazard. Read the MSDS, and follow the handling instructions. Wear appropriate protective eyewear, clothing, and gloves. Formamide should be stored aliquoted at –20°C.

11. Buffer with EDTA may cause eye, skin, and respiratory tract irritation. Read the SDS, and follow the handling instructions. Wear appropriate protective eyewear, clothing, and gloves.

12. Poly-cytosine tracts surrounding np 16189 in HV1 with length heteroplasmy generate blurred sequence readings beyond this region (see Fig. 3). The use of internal primers (L16209 and H16164) provides readings to cover the complete region edition.

13. In the case of primer L-048, the "clean" sequence starts beyond np 73. For this reason, the use of primer H-285 provides additional data to confirm the sequence in the edited range (see Fig. 4). Also in the case of length heteroplasmy produced by multiple insertions in the poly-cytosine tract around np 309, the use of primer H285 serves to read the segment beyond this region with a reverse primer.

14. Although manufacturer protocol recommends a final volume of 20 μL, the premix volume reduction does not affect negatively the quality of the EPG.

15. In a Gene-Amp® PCR System 9700 Thermal Cycler (Applied Biosystems), the program installed is Big Dye™:

    1 min at 96°C, 1 min at 56°C, and 4 min at 72°C for 25 cycles.

    If using other models, slow the ramping time.

16. Purification of extension products must be carried out in the next 24 h.
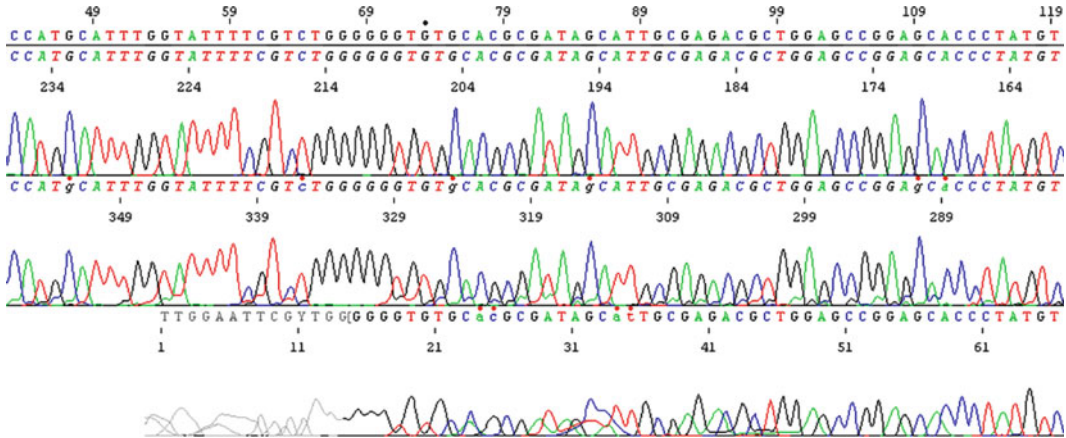
Fig. 4. EPG showing sequences generated by primers H285, H408, and L48 from *top* to *bottom.*

17. If purification is being performed with DyeEx columns or other gel filtration devices, special care must be taken to avoid touching the gel surface and to apply the sample in the center of the gel bed. Otherwise the sample can flow down the sides of the well. In this case the EPG will appear as a noisy, low-intensity sequence.

18. Speed Vac centrifuge is suitable for drying sequencing reactions. The drying time depends on the number of tubes. 40–50 min is sufficient for 20–24 tubes. Do not over-dry.

    If samples are not to be run in the sequencer, keep them in the freezer protected from light.

19. If the samples have been purified on DyeEx 96, the addition of formamide may be performed on a robotic platform. The product may be then transferred by the robotic arm to the electrophoresis sample plate avoiding sample transference-associated errors. In this case, several steps of formamide aspiration-dispense must be included in the program.

20. In the analysis protocol, the Mixed Bases Settings are 25% by default. This means that a secondary peak will be detected if the peak high is at least 25% of the main peak. This must be considered for heteroplasmy detection. In sequences with poor signal-to-noise ratio, the ability to distinguish between "true" point heteroplasmies from noisy peaks will be affected.

21. Visual inspection of EPGs sometimes shows noisy, bad quality sequences that have been "approved" by the software. One of the reasons is the use of contaminated or degraded primers. Repeat the sequencing reaction with a new primer aliquot. It should also be noted that as a consequence of its particular characteristics, not all the primers behave in the same way.
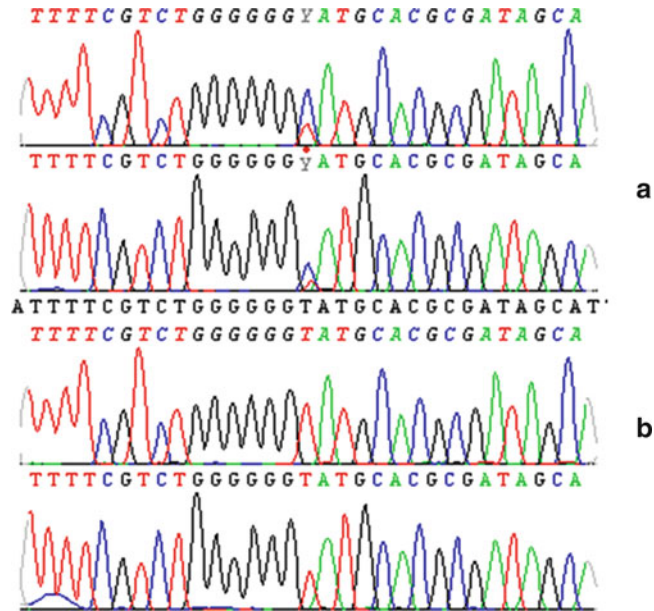
Fig. 5. EPG showing point heteroplasmy at np72 in muscle (**a**) but not in reference sample coming from a son (**b**). Haplotype (16024-576): 263G 315.1C (son) and 72Y 263G 315.1C (mother). The analysis of nuclear DNA in the trio father–mother–son (17 loci) confirmed the mother–child relationship.

Primers H285, L16555, and L15997 are particularly prone to generate good quality (clean, well-balanced, low-noise) sequences when compared with the others.

22. Especially, mtDNA from muscle tissue is prone to point heteroplasmy, mainly in aged individuals. For this reason, in samples coming from human remains consisting of muscle cells, it is very frequent to find heteroplasmic positions that are absent in the reference sample of the relative (see Fig. 5).

References

1. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 74(12):5463–7.

2. Wilson MR, DiZinno JA, Polansky D, Repogle J, Budowle B (1995) Validation of mitochondrial DNA sequencing for forensic casework analysis. Int J Legal Med 108:68–74.

3. Parson W, Bandelt H-J (2007) Extended guidelines for mtDNA typing of population data in forensic science. Forensic Science International: Genetics 1(1):13–9.

4. Lutz S, Weisser H-J, Heizmann J, Pollak S (1996) mtDNA as a tool for identification of human remains. Int J Legal Med 109: 205–9.

5. Parsons T J. Mitochondrial DNA Genome Sequencing and SNP Assay Development for increased Power of Discrimination. http://www.ncjrs.gov/pdffiles1/nij/grants/213502.pdf.

6. Parsons TJ, Coble MD (2001) Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. Croatian Med J 42: 304–9.

7. Coble MD, Just RS, O'Callaghan JE, Letmanyi IH, Peterson CT, Parsons TJ (2004) Single nucleotide polymorphisms over the entire mtDNA genome that increase the forensic power of testing in Caucasians. Int J Legal Med (2004) 118:137–46.

8. B. Quintans, V. Alvarez-Iglesias, A. Salas, C. Phillips, M.V. Lareu, A. Carracedo (2004) Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. Forensic Sci. Int. 140:251–7.

9. 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.

10. Berger C, Parson W (2009) Mini-midi-mito: Adapting the amplification and sequencing strategy of mtDNA to the degradation state of crime scene samples. Forensic Science International: Genetics 3:149–153.

11. Qiaquick spin Handbook. QIAquick PCR Purification Kit. QIAGEN 1051746.

12. 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. Nature Genetics 23(2):147 [Gen Bank NC_012920].

13. Big-Dye Terminator v3.1 Cycle Sequencing Kit. Applied BioSystems. Part Number 4337035.

14. Dye Ex™ Handbook. DyeEx 2.0 Spin Kit and DyeEx 96 Kit. QIAGEN. Ref 1019971.

15. Carracedo A, Bär W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland M, Tully G, Wilson M (2000) DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int. 110(2):79–85.

16. Tully G, Bär W, Brinkmann B, Carracedo A, Gill P, Morling N, Parson W, Schneider P (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.

17. Bandelt HJ, Lahermo P, Richards M, Macaulay V (2001) Detecting errors in mtDNA data by phylogenetic analysis. Int. J. Legal Med 115:64–9.

18. Röhl A, Brinkmann B, Forster L, Forster P (2001) An annotated mtDNA database. Int J Legal Med 115:29–39.

19. Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V (2002) The fingerprint of phantom mutations in mitochondrial DNA data. Am J Hum Genet 71:1150–60.

20. Yao YG, Bravi CM, Bandelt HJ (2004) A call for mtDNA data quality control in forensic science. Forensic Science International 141(1):1–6.

21. Brandstätter A, Sänger T, Lutz-Bonengel S, Parson W, Béraud-Colomb E, Wen B, Kong QP, Bravi CM, Bandelt HJ (2005) Phantom mutation hotspots in human mitochondrial DNA. Electrophoresis 26(18):3414–29.

22. Parson W (2006) The Art of Reading Sequence Electropherograms. Annals of Human Genetics 71:276–78.

23. Bandelt HJ, Salas A, Bravi CM (2006) What is a 'novel' mtDNA mutation--and does 'novelty' really matter? J Hum Genet 51(12): 1073–82.

24. Irwin JA, Saunier JL, Niederstatter H, Strouss KM, Sturk KA,. Diegoli TM, Brandstatter A, Parson W, Parsons TJ (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 68:516–27.

25. Salas A, Bandelt HJ, Macaulay V, Richards MB (2007) Phylogeographic investigations: the role of trees in forensic genetics. Forensic Sci Int 168(1):1–13.

26. Malyarchuk BA, Rogozin IB, Berikov VB, Derenko MV (2002) Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region. Hum Genet 111(1):46–53.

27. Del Bo R, Crimi M, Sciacco M, Malferrari G, Bordoni A, Napoli L, Prelle A, Biunno I, Moggio M, Bresolin N, Scarlato G, Pietro Comi G (2003) High mutational burden in the mtDNA control region from aged muscles: a single-fiber study. Neurobiol Aging 24(6):829–38.

28. van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30(2):E386-E394. http://www.phylotree.org.

29. Bandelt HJ, Parson W: Consistent treatment of length variants in the human mtDNA control region: a reappraisal. Int J Legal Med 2008, 122:11–21.

30. Parson W, Dür A (2007) EMOP - a forensic mtDNA database. Forensic Sci Int Genet 1:88–92.

31. Turchi C, Buscemi L, Previdere C, Grignani P, Brandstätter A, Achilli A, Parson W, Tagliabracci A (2008) F.I.Group Italian mitochondrial DNA database: results of a collaborative exercise and proficiency testing. Int J Legal Med 122:199–204.

32. Prieto et al (2010) The GHEP–EMPOP collaboration on mtDNA population data - A new resource for forensic casework. Forensic Sci Int Genet 5:146–51.

Capillary Electrophoresis of Human mtDNA Control Region Sequences from Highly Degraded Samples Using Short mtDNA Amplicons

Odile M. Loreille and Jodi A. Irwin

Abstract

The forensic applications of mtDNA sequencing center primarily on samples that are either highly degraded or contain little or no nuclear DNA, since the testing of these sample types is often unsuccessful with more widely used nuclear STR profiling assays. In these cases, sequence data from the noncoding mtDNA control region are targeted due to its high variability. However, the ease of authentic DNA recovery and the strategy used for recovery depend strictly on the quality of the sample. In this chapter, we will cover mitochondrial DNA sequencing procedures for short mtDNA amplicons which range in size from 100 to 350 bp. Generally speaking, amplicons of this size are required only for the most degraded specimens, and the protocols described here have been specifically developed for recalcitrant human skeletal remains encountered during the course of a large-scale missing persons' identification effort. DNA templates from these types of specimens tend to exhibit various forms of intrastrand damage that, in turn, manifest as artifacts in the sequence data. Because these artifacts are not generally observed among sequence data from pristine templates, we address the particular data idiosyncrasies that warrant additional scrutiny. Although this chapter will primarily highlight this particular application, the basic experimental parameters and data considerations should easily extend to other applications and/or sample types. The protocols described here have been deliberately designed to produce raw sequence electropherograms and final mtDNA profiles that adhere to the strictest forensic guidelines in terms of overall data quality.

Key words: Mitochondrial DNA, Forensic genetics, DNA identification, Control region, DNA sequencing

1. Introduction

Despite the interest of the forensic community in recovering DNA data from the smallest traces of evidentiary material, mtDNA testing is not routinely employed by most forensic DNA laboratories. The community has instead, and understandably, favored the use

of multiple autosomal short tandem repeats (STRs) because they provide much higher discriminatory power than data from the mitochondrial genome (1). And while the fairly recent developments of mini-STR multiplexes and higher efficiency STR kits have extended the application of STR testing to more degraded samples (2–9), there remain many specimens for which nuclear DNA is too degraded or too scarce to yield forensically informative STR profiles, no matter how small the amplicon or how aggressive the amplification parameters. MtDNA testing is often the only remaining DNA-based testing option for samples such as shed hairs, fingernails, and severely degraded skeletal elements, due to the relative abundance of intact mtDNA (present in several thousand copies per cell) as compared to nuclear DNA in these types of specimens (10–18). MtDNA data are also quite useful in particular forensic investigations because the molecule is maternally inherited without recombination (19–21). As a result, all maternal relatives possess identical mtDNA haplotypes in the absence of mutation and thus even distant relatives can provide reference material for particular mtDNA comparisons (16). This is extremely beneficial in those cases for which direct references, or references from immediate family members, are unavailable for standard autosomal STR testing. Even in those cases for which mtDNA is *not* the only potential source of genetic information, mitochondrial data can often be effectively used to increase (or decrease) the statistical support for a given hypothesis (e.g., if the only other DNA evidence is a common Y-haplotype). Whatever the case may be, as DNA typing methods are increasingly employed on evermore small, damaged, and degraded specimens, and as DNA-based missing persons, cold case, and mass fatality initiatives expand, more and more forensic laboratories will turn to mtDNA testing for its obvious benefits with the most difficult samples (22–28).

Forensic laboratories already performing large-scale mtDNA testing of severely degraded remains owe much of their progress to developments in the field of ancient DNA, and particularly, its early successes and failures. The mitochondrial genome has been the focus of various ancient DNA studies for nearly 30 years, and it was the early years of molecular archeology research that clearly demonstrated the difficulty of recovering authentic ancient DNA. Few projects were successful, and the results of many early studies reporting DNA recovery from million year-old specimens turned out to be contamination artifacts that were only identified following publication (29–33). While such lessons learned are neither surprising nor unique in the history of science, it took several years for the scientific community to realize that ancient mtDNA testing was, among other things: (1) highly prone to contamination (much more so than STR testing) and (2) complicated by miscoding lesions caused by postmortem DNA damage and degradation (34–38). As a result, strict guidelines are now in place for laboratory set-up, experimental protocols, replicate testing, and data analysis (39, 40).

From the standpoint of laboratory testing for routine forensic mtDNA casework, the noncoding control region (CR), and its two hypervariable segments (HVS) in particular, are regularly targeted because of their high variability (41, 42). These approximately 600 bases have the highest average substitution rate in the mitochondrial genome, and thus present the greatest opportunity for interindividual differentiation while minimizing data generation effort. The chance that any two randomly sampled individuals match in this region is generally less than 1%, but depends on the population (16).

The size of the mtDNA amplicons that can be targeted in any given case is strictly correlated to the state of preservation of the DNA and, in turn, the sample. Nucleic acids generally survive better in cold environments because of the lower activity of enzymes and microorganisms involved in the first stages of degradation (autolysis and putrefaction). On the other hand, specimens subjected to hot and/or humid climates where both endogenous and exogenous microbial activity is high, or samples exposed to hydrogen, oxygen, and/or ultra-violets, tend to experience greater nucleic acid damage and degradation (43). In particular instances, DNA damage may also be caused by chemical treatment of the remains—as a result of various mortuary practices (44) or perhaps as a result of deliberate efforts to destroy any evidence of the human remains (45, 46). Regardless of the precise postmortem processes at work, nucleic acids present in the most recalcitrant specimens are nearly always highly fragmented and in low quantity, precluding the amplification of mtDNA fragments of decent size. Given this, along with the large variability in DNA quantity and quality on a sample-by-sample basis, it is generally helpful to estimate the average length of the surviving endogenous fragments prior to selection of amplification primers. A few specific quantitative PCR assays which target mitochondrial DNA fragments of various sizes have proven quite useful for this purpose (47–50). In our laboratory, the qPCR results dictate whether primers that target fragments between 211 and 330 bp (Primer Sets: PS, see Tables 1 and 2) or primers producing smaller amplicons ranging from 124 to 189 bp should be used (Mini Primer Sets: MPS, see Tables 1 and 2). Any attempt to amplify fragments that are longer than the average surviving fragment in the extract generally results in amplification failure or, if a very small number of molecules actually permit the reaction to start, sequences with a high rate of error due to damaged starting templates and, in some rare cases, jumping PCR products (51). The protocol described here has been optimized for short mtDNA amplicons derived from severely damaged and degraded human skeletal remains. It is based on fluorescent dye-terminator chemistries and capillary-based electrophoresis.

Table 1
Primer sets (PS) and miniprimer sets (MPS) combinations

Primers sets (+amplicon size in brackets)
  Primer Set 1 **F15989/R16251*** [263] or F15971/R16258 [288]
  Primer Set 2 **F16190/R16410**-M19 [221] or F16190/R16400 [211]
  Primer Set 3 **F15/R285** [271] or F15/R270 [256] or F29/R270 [242] or F29/R274 [246]
  Primer Set 4 F140/R389 [250] or **F155/R389** [235] or F155/R381 [227] or F155/R484 [330]
  Primer Set 5 F16363/R16 [223]

Miniprimer sets
  MPS1A F15989/R16158 [170]
  MPS1B **F16112/R16251*** [140] or F16112/R16237 [126]
  MPS2A F16190/R16322 [133]
  MPS2B F16222/R16410-M19 [189] or **F16268/R16410**-M19 [143] or F16222/R16400 [179]
  or F16268/R16400 [133]
  MPS3A F34/R159 [126]
  MPS3B F109/R240 [132]
  MPS4A F151/R292 [142]
  MPS4B F220/R389 [170]
  MPS5A F16363/R16509 [147]

Miniprimer sets when variable region testing is necessary
  MVR1 F16450/R16 [136]
  MVR2 F403/R569 [167]

For amplification conditions, see ref (59). The pairs in bold are the standard combinations that are most commonly used. *For higher data quality, R16237 should replace R16251 for sequencing.

2. Materials

2.1. Presequencing Components

1. Agarose gel or FlashGel (Cambrex, East Rutherford, NJ, USA), quantification ladder and dye.
2. Exonuclease 1 (Exo. USB Corporation, Cleveland, OH).
3. Shrimp Alkaline Phosphatase (SAP. USB Corporation, Cleveland, OH).

   Or

4. MinElute PCR Purification Kit and/or QiAquick PCR purification kit (Qiagen, Valencia, CA).

2.2. Cycle Sequencing Components

1. BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) (see Note 1).
2. dGTP BigDye® Terminator Cycle Sequencing Kit (Applied Biosystems) (see Note 2).

Table 2
Primer sequences

Primer sequences
F15971 5′-TTA ACT CCA CCA TTA GCA CC-3′
F15989 5′-CCC AAA GCT AAG ATT CTA AT-3′
F16112 5′-CAC CAT GAA TAT TGT ACG GT-3′
F16190 5′-CCC CAT GCT TAC AAG CAA GT-3′
F16222 5′-CCT CAA CTA TCA CAC ATC-3′
F16268 5′-CAC TAG GAT ACC AAC AAA CC-3′
F16363 5′-CCC CAT GGA TGA CCC CCC-3′
F16450 5′-GCT CCG GGC CCA TAA CAC TTG-3′
R16158 5′-TAC TAC AGG TGG TCA AGT AT-3′
R16237 5′-TGT GTG ATA GTT GAG GGT TG-3′
R16251 5′-GGA GTT GCA GTT GAT GT-3′
R16258 5′-TGG CTT TGG AGT TGC AGT TG-3′
R16322 5′-TGG CTT TAT GTA CTA TGT AC-3′
R16400 5′-GTC AAG GGA CCC CTA TCT GA-3′
R16410-M19 5′-GAG GAT GGT GGT CAA GGG A-3′
R16509 5′-AGG AAC CAG ATG TCG GAT AC-3′
F15 5′-CAC CCT ATT AAC CAC TCA CG-5′
F29 5′-CTC ACG GGA GCT CTC CAT GC-3′
F34 5′-GGG AGC TCT CCA TGC ATT TGG TA-3′
F109 5′-GCA CCC TAT GTC GCA GTA TCT GTC-3′
F140 5′-CCT GCC TCA TCC TAT TAT TTA-3′
F151 5′-CTA TTA TTT ATC GCA CCT-3′
F155 5′-TAT TTA TCG CAC CTA CGT TC-3′
F220 5′-TGC TTG TAG GAC ATA ATA AT-3′
F403 5′-TCT TTT GGC GGT ATG CAC TTT-3′
R16 5′-TGA TAG ACC TGT GAT CCA TCG TGA-3′
R159 5′-AAA TAA TAG GAT GAG GCA GGA ATC-3′
R240 5′-TAT TAT TAT GTC CTA CAA GCA-3′
R270 5′-TGG AAA GTG GCT GTG CAG AC-3′
R274 5′-TGT GTG GAA AGT GGC TGT GC-3′

Table 2
(continued)

Primer sequences

R285 5′-GTT ATG ATG TCT GTG TGG AA-3′
R292 5′-ATT TTT TGT TAT GAT GTC T-3′
R381 5′-GCT GGT GTT AGG GTT CTT TG-3′
R389 5′-CTG GTT AGG CTG GTG TTA GG-3′
R484 5′-TGA GAT TAG TAG TAT GGG AG-3′
R569 5′-GGT GTG TTT GGG GTT TGG TTG-3′
R599 5′-TTG AGG AGG TAA GCT ACA TA-3′

3. HPLC purified primers at 10 μM; Sequence data from both forward and reverse primers are desirable, but not always possible to obtain. See Table 2 for primer sequences and Note 3.

4. Big Dye dilution buffer (400 mM Tris, 10 mM MgCl$_2$, pH 9.0) (see Note 4).

5. Sterile, deionized water.

6. MicroAmp® Optical 96-Well Reaction Plate (Applied Biosystems) or microtubes.

7. Ice is recommended to reduce nonspecific priming.

8. Cycle PCR sealers (Fisher scientific, Pittsburgh, PA) or optical caps (Applied Biosystems).

9. Tips, filtered/aerosol-resistant.

2.3. Sequence Reaction Purification and Loading

1. AGTC gel filtration purification columns or 96-well blocks (Edge Biosystems, Gaithersburg, MD, USA).

2. Highly deionized (Hi-Di) Formamide (Applied Biosystems).

3. 96-Well Plate Septa (Applied Biosystems).

4. 96-Well Plate Base (Applied Biosystems).

5. 96-Well Plate Retainer (Applied Biosystems).

6. POP-6™ Polymer for 3130/3130*x*l (Applied Biosystems).

7. Running Buffer, 10× (Applied Biosystems) (see Note 5).

2.4. Equipment

1. Gel Electrophoresis apparatus.

2. Thermo cycler.

3. Vortex.

4. Centrifuge with tray carriage.

5. Centrifugal evaporator (Speed Vac evaporator).

6. ABI 310, 3100, or 3130xl Genetic Analyzer (Applied Biosystems) (see Note 6).

7. Electronic pipettes are strongly recommended.

3. Methods

In terms of degraded evidence testing, the largest hurdles to data acquisition have largely been surmounted once the DNA sequencing stage is reached. The testing of evidence material is guided by the original quantity and quality of the specimen, which also dictates the quantity and quality of the DNA. Therefore, the laboratory steps that precede DNA sequencing are critical to the recovery of authentic, high-quality sequence data that can be reported in forensic cases. In order to avoid contamination at these critical steps, a few specific guidelines should be followed when processing evidentiary specimens (39, 40). Laboratories where samples are extracted and amplification reactions are prepared should be physically separated from laboratories where amplified DNA is handled. Each laboratory should have dedicated equipment, reagents, and supplies so that transfers between pre- and post-PCR laboratories are avoided. Experimental controls should be included at all steps of laboratory processing since extraction blanks and negative PCR controls are critical for monitoring contamination. Replicate testing should be conducted on all samples, starting at the extraction step if possible, to ensure consistent and reproducible results; and finally, the data should be critically evaluated to ensure that they logically "make sense" (for review, see ref. (39)).

3.1. PCR Product Quantification

Evaluate the amplification results on a 2% agarose gel or a FlashGel and assess DNA quantity using either a quantification ladder or a spectrophotometer. For amplicons between 100 and 200 bp, use 1–3 ng of PCR product for sequencing. For amplicons between 200 and 400 bp, use 3–5 ng of PCR product as recommended by the manufacturer (see Note 7).

3.2. PCR Product Purification

Purify successful reactions by adding 10 μL of Exo and 5 μL of SAP to the reactions. Place the tubes in the thermo cycler for 45 min at 37°C, 20 min at 80°C, followed by a final hold at 4°C (see Note 8). If 2 μL of amplified product contain significantly less than 1 ng of DNA, use the MinElute kit following the manufacturer's instructions

instead of the Exo and SAP to purify the reactions, and elute your DNA in a volume smaller than your reaction volume to concentrate it. The MinElute or the QiAquick PCR purification kits should also be used when many primer dimers are present in the reactions.

3.3. Cycle Sequencing Reaction

The use of electronic pipettes is recommended for the set-up of many sequencing reactions, to save time and decrease pipetting errors. In a 96-well reaction plate or, if only a small number of reactions are being prepared, in carefully labeled microtubes: add 2 μL of one sequencing primer (10 μM) in each well, along with 1 or 2 μL of purified DNA. Prepare a sequencing reaction master mix that includes, per sample: 8 or 9 μL of deionized water, 6 μL of big dye dilution buffer, 1.5 μL of big dye, 0.5 μL of dGTP. Distribute 16 or 17 μL of the master mix in each well to have a final volume of 20 μL per well. Add the plate sealer and briefly vortex the plate. Spin it and place it in a thermo cycler for 25 cycles of 15 s at 96°C, 5 s at 50°C, 2 min at 60°C followed by a rapid thermal ramp to 4°C.

3.4. Sequence Purification

Purify sequencing reactions with AGTC gel filtration purification columns according to the manufacturer recommendations and then dry the samples in a centrifugal vacuum evaporator for 60 min or until dry (caution: do not heat and do not overdry). Resuspend the samples in 10 μL of Hi-Di Formamide, add a plate septum, and vortex. After a brief centrifugation, add the 3130 base plate and retainer, and finally place the plate in the sequencer. A heat denaturation step is not necessary.

3.5. Electrophoresis Parameters and Analysis

The Standard RapidSeq36_POP6 Run Module (Instrument Protocol) from Applied Biosystems can be customized slightly to accommodate the short injection times used for small amplicon sequencing. The specific injection time will vary according to the DNA concentration and amplicon size, but generally speaking, very short injection times of 3–7 s should be used. Short injections will help prevent overloading and the resultant high background that is caused by off-scale data and spectral pull-up.

For sequence analysis, the standard RRSeqAnalysisPOP6 module/protocol should be used. As with the run module, the analysis module will need to be slightly modified to accommodate the short fragments. The start and, more importantly, stop points of the data analysis (which correspond to the start and stop points of the instrument scans that will be analyzed by the software) should be properly set to accommodate the short fragment sizes. In other words, take a look at the raw data to determine a scan number just beyond the end point of the longest sequence fragments, and then use that as the stop point for the data analysis. This will limit

the data analysis to a shorter range, yet also capture the length of all sequences (see Notes 9 and 10).

3.6. Data Review and Establishment of Authenticity

Details of standard mitochondrial DNA data review are covered in detail in Chapters 19 and 21, so here we will focus only on data features/characteristics that are specific to damaged and degraded DNA templates.

In their publication (39), Gilbert et al. explain that the criteria of authenticity described in various reviews on ancient DNA are "intended to assist in determining the authenticity of a study, but [..] cannot replace a crucial consideration of the problem." The idea is that, the data must always be reviewed carefully to determine whether or not the results obtained make sense. Below, we list a few "sense"-based strategies to help establish data authenticity. In addition, we present some particular scenarios that warrant further scrutiny of the data and should be treated with caution.

1. The best way to confirm the authenticity of a sequence is by replicating the amplifications several times from at least two independent extracts. In our laboratory, different scientists perform the two DNA extractions as an additional precautionary measure. In order to avoid false positives and/or contamination from contemporary molecules that could have potentially penetrated a particular sample (a porous bone for example), it is strongly recommended that different skeletal elements be used for the independent extractions (for example, a bone and a tooth, or a humerus and femur).

2. When possible, try to perform 3 or 4 replicate amplifications of a given DNA extract in a single PCR set-up. If all of the reactions containing DNA show PCR product, even if the bands of the agarose gel are faint, you can consider the results encouraging. If, on the other hand, you only detect PCR product in a single amplification of the many replicates, then this should raise a red flag. Independent amplifications should be performed anyway, but multiple replicates in a single set-up provide a much better picture of reproducibility than a single amplification does.

3. Electropherograms should be reviewed base-by-base, in order to identify any peculiarities. A mixed base can have different origins. If a position shows mixtures in all the replicate amplification products and the PCR negative controls are negative (see Fig. 1), it is very likely that either the DNA extract is contaminated or the PCR has been contaminated by a quantity of DNA so small that it is insufficient to initiate an amplification in the controls, but when combined with a few molecules of authentic DNA, will amplify. This is known as the carrier effect (52, 53); and while it can be caused by a number of things, past amplification products are known to be one source (54). Contamination is also very likely when the mixed bases occur
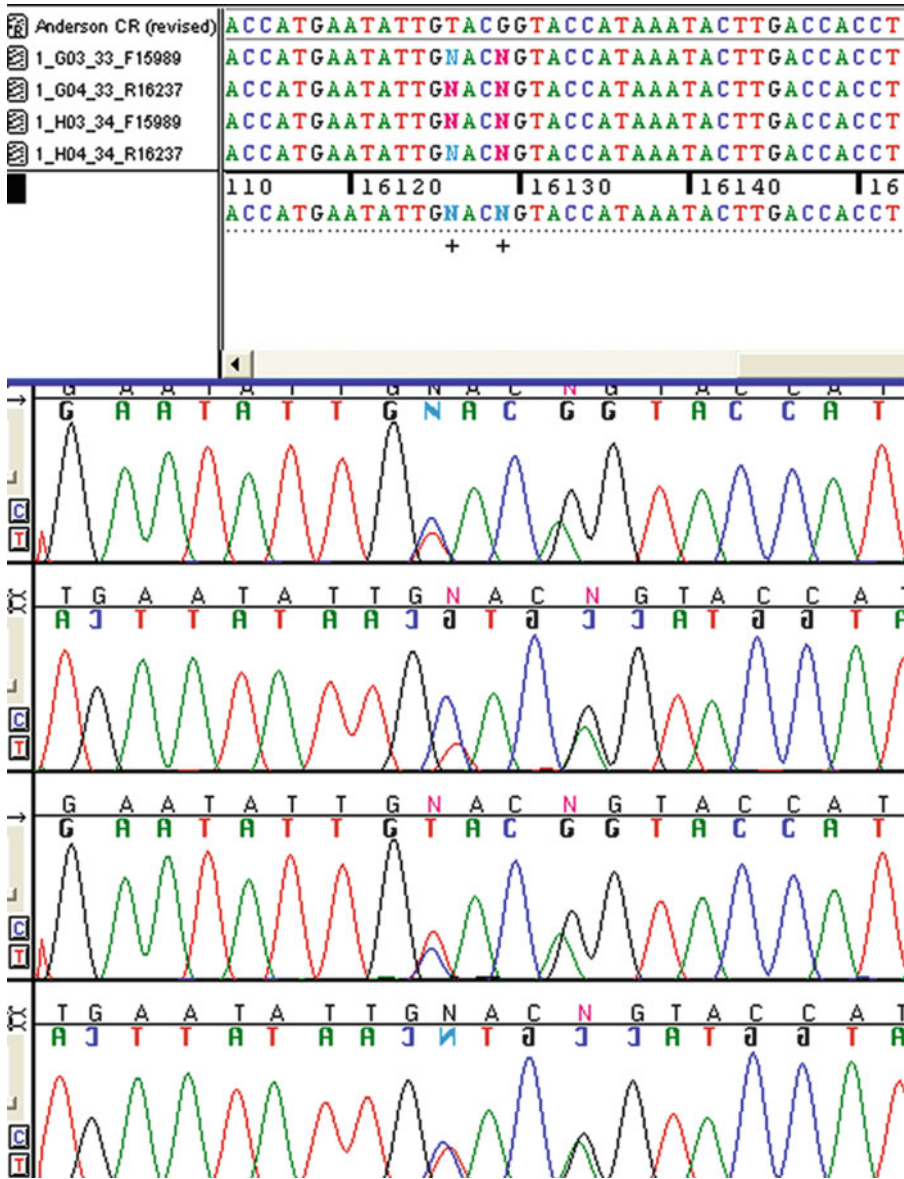
Fig. 1. Example of a contaminated sequence. Mixtures are observed at variable and phylogenetically informative positions 16126 and 16129 and the respective forward and reverse sequences represent independent amplifications.

at sites that are phylogenetically informative, or at sites generally known to vary in the human mtDNA phylogeny (55, 56). On the other hand, when the mixed sites seem random among replicate amplifications and sequences (see Fig. 2), and they occur at positions that are phylogenetically uninformative and/ or rarely vary, it is likely that the mixed bases are the result of Taq polymerase errors induced by damaged DNA templates. The number of polymerase misincorporations is strictly linked to the quantity and severity of DNA damage.
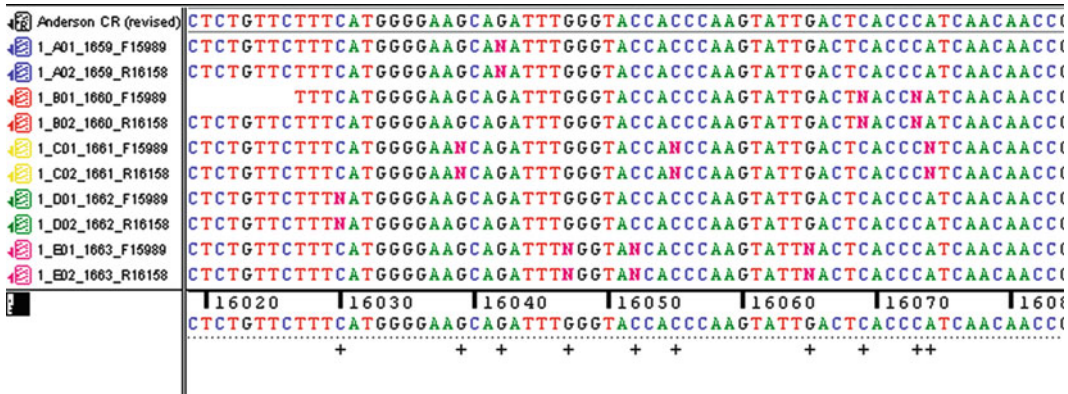
Fig. 2. Example of a sequence obtained from an aggressively embalmed sample that shows a high quantity of irresolvable bases due to DNA damage. The ten sequences shown represent the forward and reverse sequences from five separate amplicons. Each sequence pair (F and R) exhibiting the same constellation of mixed bases originates from a different amplification replicate. Note how the mixed bases are not reproduced among replicates and tend to occur at positions that rarely, or never, vary in the human mtDNA phylogeny.
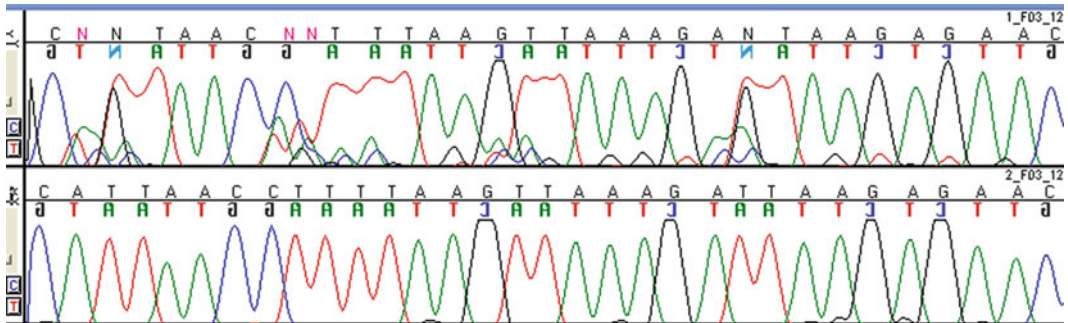
Fig. 3. Example of an overblown sequence (*top*) compared to the same sequencing reaction after a shorter 3130 injection time (*bottom*). The RFU values from the top sequence range between 6360 and 2851. With a shorter injection time, the background noise disappeared and the RFUs fell down to between 229 and 77.

4. A short amplicon will often produce sequences containing a lot of background noise and sequence data artifacts. Sometimes, however, the background noise is simply caused by too much sequencing product. In these cases, rerunning the sample with a shorter injection time will generally solve the problem (see Fig. 3).

5. Although extremely rare when using human-specific mtDNA control region primers, a sequence may be encountered on occasion that exhibits an unusually large number of differences from the rCRS, with many of those differences occurring at sites that are generally considered invariant in the human mtDNA phylogeny. These sequences are by-and-large quite obvious and should be searched against GenBank. In such cases, it is highly likely that a pseudogene has been inadvertently amplified instead of the mtDNA target (57).
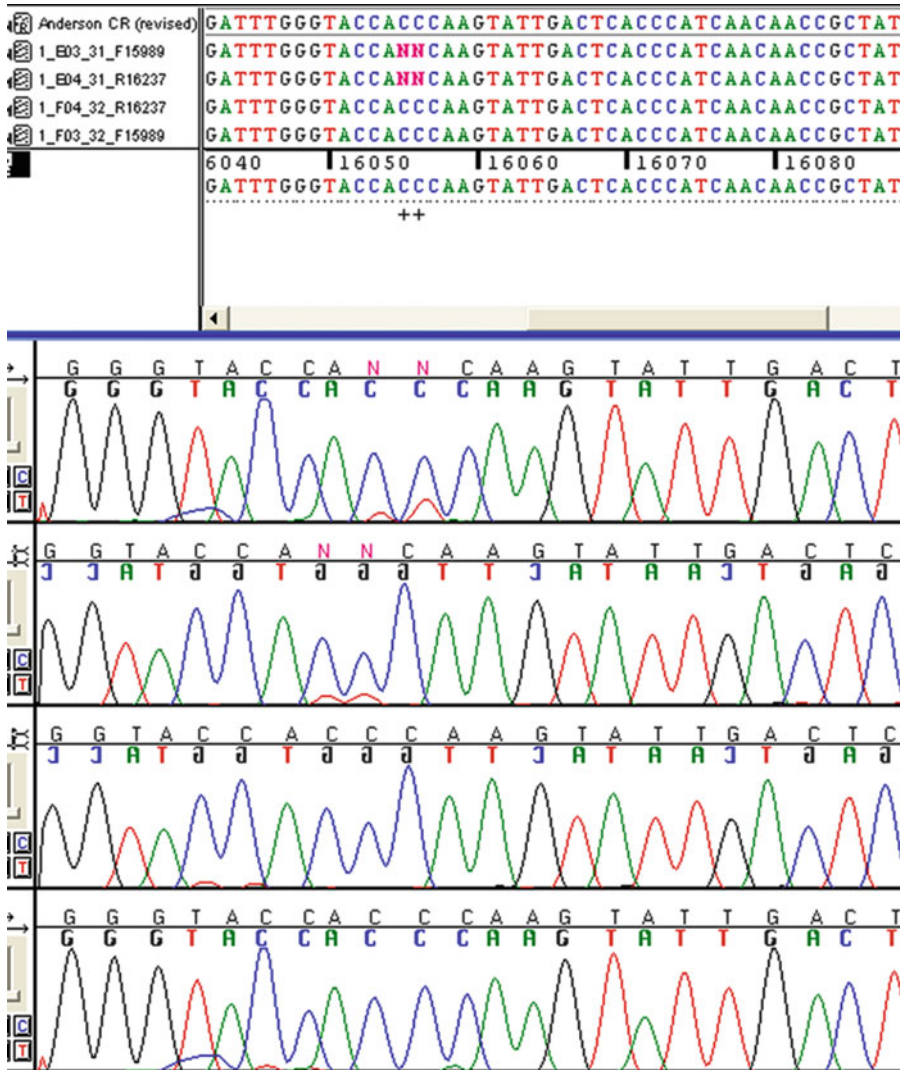
Fig. 4. Example showing deamination. T peaks can be observed under the C-stretch. Of the four sequences shown, the two upper sequences represent the forward and reverse primers from one amplicon and the two lower sequences represent the same primers from a replicate amplification. Deamination, like other damage, tends to be difficult to replicate; and that is the case here. Deamination products in the two amplicons occur at different positions (16055–16056 in the upper sequences, 16052–16053 in the lower sequences).

6. Deamination of, in particular, adenine and cytosine is a common form of DNA modification in highly degraded specimens (35). This type of damage produces sequence data artifacts that manifest as mixed bases of Cs and Ts or, the complementary Gs and As. This is often observed in C-stretches, where small red peaks (T) are found under blue peaks (C) (see Fig. 4). However, like other forms of DNA damage, deamination can generally be easily identified by its random occurrence at

different positions in different PCR products from the same DNA extract. This is due to a limited number of starting molecules and the stochastic amplification of one template over another during the early PCR cycles, which leads to the overrepresentation of one template over any other in a given amplification. It is important to note, however, that in all cases of DNA damage, the background haplotype/haplogroup should remain consistent across all amplifications. In other words, any phylogenetically informative or authentic DNA variation should remain stable across the replicates (unless, by pure coincidence, the damage occurred at one of those sites).

4. Notes

1. The BigDye® Terminator v1.1 Cycle Sequencing Kit has been optimized to deal with base calling directly adjacent to the primer. We tried both version 1.1 and v3.1 of these kits and our experience confirmed that the BigDye® Terminator v3.1 Cycle Sequencing Kit does not perform as well as the v1.1 with short amplicon sequencing.

2. The dGTP BigDye Terminator kit is optimized for G and GT-rich templates and generally improves the quality of the reverse (H-strand) sequencing reactions—particularly in poly-G regions. The reaction mix of this kit contains a mixture of dGTPs, dATPs, dTTPs, and dCTPs. However, on a 3100 or a 3130 Genetic Analyzer, the use of the dGTP kit alone may cause peak compression issues. Therefore, for these instruments, it is best to use the BigDye® Terminator v1.1 Cycle Sequencing reaction mix (that contains inosine, dITP, instead of dGTP) spiked with the dGTP BigDye® Terminator Cycle Sequencing reaction mix as described in Subheading 3.3.

3. For amplicons smaller than 100 bp, regular Sanger sequencing will provide low quality data. However, an option in these cases is the use of poly-tailed sequencing primers to artificially lengthen the sequenced fragments to >100–125 bp—the size generally necessary for good quality sequence data with these instruments. Binladen et al. recommend the addition of a neutral 40 bp sequence (5′-AACTGACTAAACTAGGTGCCACG-TCGTGAAAGTCTGACAA-3′) and additional 5′ poly(C) tail to the 5′ end of the basic sequencing primer (58). Our preliminary work using primers + the 40 bp sequence (but no poly-C tail) gave very promising results (data not shown).

4. The dilution sequencing buffer can be purchased from Applied Biosystems. To prepare it yourself, dissolve 48.4 g Tris base and 2.0 g $MgCl_2$ in 800 mL dH_2O. Adjust to pH 9.0 at room tem-

perature by adding drops of concentrated HCl. Adjust the final volume to 1 L with dH$_2$O and mix thoroughly. Filter the buffer. Caution: Hydrochloric acid (HCl) causes severe burns and is irritating to the eyes. When preparing this reagent, use a fume hood and avoid inhalation and contact with the skin. Wear a lab coat, gloves, and protective eyewear when handling.

5. 10× running buffer can be purchased from Amresco (Solon, OH) and used instead of the Applied Biosystems buffer to reduce costs.

6. The 3730 DNA Analyzer of Applied Biosystems requires the use of the POP-7™ Polymer. Our experience has shown that this instrument/polymer combination did not provide the resolution necessary for very short fragments.

7. The frequency of nonspecific amplification products is significantly higher when DNA templates are very degraded, and/or sample extracts are contaminated by large quantities of micro-organismal DNA. If the agarose gels show multiple bands in a single lane, several options are available to overcome this problem. (1) Try another amplification using more stringent conditions (often, a lower concentration of magnesium chloride and/or higher annealing temperature is sufficient), (2) cut and purify the desired PCR product from the agarose gel, or (3) use internal primers during the cycle sequencing step. Alternatively, different PCR primers can often be used to produce a more specific amplification product and avoid nonspecific primer binding. Multiple PCR primer options, along with various internal sequencing primers for any given amplicon (if the amplicon is large enough), can generally mitigate most problems related to nonspecific amplification (see Table 1).

8. Due to the small total volume pipetted (15 μL) and the high viscosity of Exo and SAP, prepare a master mix for at least 2 or 3 extra reactions.

9. If the signal is too strong, try rerunning the plate with a shorter injection time and/or diluting the sequences with additional HiDi formamide. If this isn't enough, perform a new cycle sequencing reaction with less, or diluted, PCR product.

10. Data analysis extending beyond the end of the sequenced fragment can sometimes yield nonsensical electropherograms, with the fluorescent signal seemingly extrapolated across the entire analysis frame. These e-grams often exhibit extremely broad peaks spanning multiple base positions, resulting in a sequence of AGTC being called as AAAGGGTTTCCC. To avoid this, the analysis start and end points can either be changed in the analysis module that is run immediately after the run module, or they can be adjusted later, on a sequence-by-sequence basis in the Sequencing analysis software.

Disclaimer

The opinions and assertions contained herein are solely those of the authors and are not to be construed as official or as views of the United States Department of Defense or the United States Department of the Army. Certain commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the United States Department of Defense or the United States Department of the Army.

Acknowledgments

References

1. Lygo, J.E., Johnson, P.E., Holdaway, D.J., Woodroffe, S., Whitaker, J.P., Clayton, T.M., et al. (1994) The validation of short tandem repeat (STR) loci for use in forensic casework. *Int J Legal Med* 107, 77–89.

2. Butler, J.M., Shen, Y., and McCord, B.R. (2003) The Development of reduced size STR amplicons as tools for analysis of degraded DNA. *J Forensic Sci* 48, 1054–1064.

3. Coble, M.D. and Butler, J.M. (2005) Characterization of New MiniSTR Loci to aid analysis of degraded DNA. *J Forensic Sci* 50, 1–11.

4. Asamura, H., Sakai, H., Kobayashi, K., Ota, M., and Fukushima, H. (2006) MiniX-STR multiplex system population study in Japan and application to degraded DNA analysis. *Int J Legal Med* 120, 174–181.

5. Asamura, H., Sakai, H., Ota, M., and Fukushima, H. (2007) MiniY-STR quadruplex systems with short amplicon lengths for analysis of degraded DNA samples. *Forensic Sci Int Genet* 1, 56–61.

6. Parsons, T.J., Huel, R., Davoren, J., Katzmarzyk, C., Milos, A., and Selmanović, A. (2007) Application of novel "mini-amplicon" STR multiplexes to high volume casework on degraded skeletal remains. *Forensic Sci Int Genet* 1, 175–179.

7. Fondevila, M., Phillips, C., Naveran, N., Fernandez, L., Cerezo, M., Salas, A., et al. (2008) Case report: identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. *Forensic Sci Int Genet* 2, 212–218.

8. Mulero, J.J., Chang, C.W., Lagacé, R.E., Wang, D.Y., Bas, J.L., McMahon, T.P., et al. (2008) Development and validation of the AmpFlSTR MiniFiler PCR Amplification Kit: a MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA. *J Forensic Sci* 53, 838–852.

9. Diegoli, T.M. and Coble, M.D. (2010) Development and characterization of two mini-X chromosomal short tandem repeat multiplexes. *Forensic Sci Int Genet* 5, 415–421.

10. Pääbo, S. (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci* 86, 1939–1943.

11. Hagelberg, E. and Clegg, J.B. (1991) Isolation and characterization of DNA from archaeological bone. *Proc Biol Sci* 244, 45–50.

12. Hochmeister, M.N., Budowle, B., Borer, U. V., Eggmann, U., Comey, C.T., and Dirnhofer, R. (1991) Typing of deoxyribonucleic acid (DNA)

extracted from compact bone from human remains. *J Forensic Sci* 36 1649–1661.

13. Wilson, M.R., Stoneking, M., Holland, M.M., DiZinno, J.A., Budowle, B. (1993) Guidelines for the Use of Mitochondrial DNA Sequencing in Forensic Science. *Crime Lab Digest* 20, 68–77.

14. Wilson, M. R., Polanskey, D., Butler, J., DiZinno, J. A., Repogle, J., and Budowle, B. (1995) Extraction, PCR amplification, and sequencing of mitochondrial DNA from human hair shafts. *Biotechniques* 18, 662–669.

15. Allen, M., Engström, A.S., Meyers, S., Handt, O., Saldeen, T., von Haeseler, A., Pääbo, S., and Gyllensten, U. (1998) Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. *J Forensic Sci* 43, 453–464.

16. Holland, M. and Parsons, T. (1999) Mitochondrial DNA sequence analysis – validation and use for forensic casework. *Forensic Sci Rev* 11, 22–50.

17. Melton, T. and Nelson, K. (2001) Forensic mitochondrial DNA analysis: two years of commercial casework experience in the United States. *Croat Med J* 42, 298–303.

18. Tully, G., Barritt, S.M., Bender, K., Brignon, E., Capelli, C., Dimo-Simonin, N., et al. (2004) Results of a collaborative study of the EDNAP group regarding mitochondrial DNA heteroplasmy and segregation in hair shafts. *Forensic Sci Int* 140, 1–11.

19. Giles, R.E., Blanc, H., Cann, H.M., and Wallace, D.C. (1980) Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci USA* 77, 6715–6719.

20. Macaulay, V., Richards, M., and Sykes, B. (1999) Mitochondrial DNA recombination-no need to panic. *Proc Biol Sci* 266, 2037–2039.

21. Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C., and Schatten, G. (1999) Ubiquitin tag for sperm mitochondria. *Nature* 402, 371–372.

22. Holland, M.M., Fisher, D.L., Mitchell, LG., Rodriquez, W.C., Canik, J.J., Merril, C.R. et al. (1993) Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. *J Forensic Sci* 38, 542–553.

23. Edson, S.M., Ross, J.P., Coble, M.D., Parson, T.J., and Barritt, S.M. (2004) Naming the dead—confronting the realities of rapid identification of degraded skeletal remains. *Forensic Sci Rev* 16, 63–90.

24. Irwin, J.A., Edson, S.M., Loreille, O., Just, R.S., Barritt, S.M., Lee, D.A. et al. (2007) DNA identification of "Earthquake McGoon" 50 years postmortem. *J Forensic Sci* 52, 1115–1118.

25. Palo, J.U., Hedman, M., Söderholm, N., Sajantila, A. (2007) Repatriation and identification of the Finnish World War II soldiers. *Croat Med J* 48, 528–535.

26. Lee, H.Y., Kim, N.Y., Park, M.J., Sim, J.E., Yang, W.I., Shin, K.J. (2010) DNA typing for the identification of old skeletal remains from Korean War victims. *J Forensic Sci* 55, 1422–1429.

27. Ríos, L., Ovejero, J.I., Prieto, J.P. (2010) Identification process in mass graves from the Spanish Civil War I. *Forensic Sci Int* 199, e27–e36.

28. Zupanic Pajnic, I., Gornjak Pogorelc, B., 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. Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., and Hofreiter, M. (2004) Genetic analyses from ancient DNA. *Ann Rev Genet* 38, 645–679.

30. Golenberg, E.M., Giannasi, D.E., Clegg, M.T., Smiley, C.J., Durbin, M., Henderson, D., and Zurawski, G. (1990) Chloroplast DNA sequence from a miocene Magnolia species. *Nature* 344, 656–658.

31. Soltis, P.S., Soltis, D.E., and Smiley, C.J. (1992) An rbcL sequence from a Miocene taxodium (bald cypress). *Proc Natl Aca Sci USA* 89, 449–451.

32. Cano, R.J., Poinar, H.N., Pieniazek, N.J., Acra, A., and Poinar, G.O. (1993) Amplification and sequencing of DNA from a 120–135 million-year-old weevil. *Nature* 363, 536–538.

33. Woodward, S.R., Weyand, N.J., and Bunnell, M. (1994) DNA sequence from Cretaceous period bone fragments. *Science* 266, 1229–1232.

34. Hansen, A., Willerslev, E., Wiuf, C., Mourier, T., and Arctander, P. (2001) Statistical evidence for miscoding lesions in ancient DNA templates. *Mol Biol Evol* 18, 262–265.

35. Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, A., and Pääbo, S. (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res* 29, 4793–4799.

36. Binladen, J., Wiuf, C., Gilbert, M.T., Bunce, M., Barnett, R., Larson, G., et al. (2006) Assessing the fidelity of ancient DNA sequences amplified from nuclear genes. *Genetics* 172, 733–741.

37. Gilbert, M.T., Binladen, J., Miller, W., Wiuf, C., Willerslev, E., Poinar, H., Carlson J.E., et al. (2007) Recharacterization of ancient

DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Res* 35, 1–10.

38. Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2010) Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res* 38, e87.

39. Gilbert, M.T., Bandelt, H.J., Hofreiter, M., and Barnes I. (2005) Assessing ancient DNA studies. *Trends Ecol and Evol* 20, 541–544.

40. Willerslev, E. and Cooper, A. (2005) Ancient DNA. *Proc Biol Sci* 272, 3–16.

41. Greenberg, B.D., Newbold, J.E., and Sugino, A. (1983) Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA, *Gene* 21, 33–49.

42. Horai, S. and Hayasaka, K. (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am J Hum Genet* 46, 828–842.

43. Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., and Pääbo, S. (2001) Ancient DNA. *Nat Rev Genet* 2, 353–359.

44. Koon, H.E., Loreille, O.M., Covington, A.D., Christensen, A.F., Parsons, T.J., and Collins, M.J. (2008) Diagnosing post-mortem treatments which inhibit DNA amplification from US MIAs buried at the Punchbowl. *Forensic Sci Int* 178, 171–177.

45. Coble, M.D., Loreille, O.M., Wadhams, M.J., Edson, S.M., Maynard, K., Meyer, C.E., et al. (2009) Mystery solved: the identification of the two missing Romanov children using DNA analysis. *PLoS One* 4, e4838.

46. Schwark, T., Heinrich, A., Preuße-Prange, A., and von Wurmb-Schwark,N. (2010) Reliable genetic identification of burnt human remains. *Forensic Sci Int Genet* 5, 393–399.

47. Alonso, A., Martín, P., Albarrán, C., García, P., García, O., et al. (2004) Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies. *Forensic Sci Int* 139, 141–149.

48. Pruvost, M. and Geigl, EM. (2004) Real-time quantitative PCR to assess the authenticity of ancient DNA amplification. *J Archaeol Sci* 31, 1191–1197.

49. Timken, M.D., Swango, K.L., Orrego, C., and Buoncristiani, M.R. (2005) A duplex real-time qPCR assay for the quantification of human nuclear and mitochondrial DNA in forensic samples: implications for quantifying DNA in degraded samples. *J Forensic Sci* 50, 1044–1060.

50. Niederstätter, H., Köchl, S., Grubwieser, P., Pavlic, M., Steinlechner, M., and Parson, W. (2007) A modular real-time PCR concept for determining the quantity and quality of human nuclear and mitochondrial DNA. *Forensic Sci Int Genet* 1, 29–34.

51. Pääbo, S., Irwin, D.,and Wilson, A. (1990) DNA damage promotes jumping between templates during enzymatic amplification. *J Biol Chem* 265, 4718–4721.

52. Cooper, A. (1994) DNA from museum specimens. In: Ancient DNA. Hermann B, Hummel S (Eds) New York: Springer-Verlag, 149–165.

53. Handt, O., Höss, M., Krings, M., Pääbo, S. (1994) Ancient DNA: methodological challenges. *Experientia 50*, 524–529.

54. Kwok, S., and Higuchi, R. (1989) Avoiding false positives with PCR. *Nature* 339, 237–238.

55. Gilbert, M.T., Willerslev, E., Hansen, A.J., Barnes, I., Rudbeck L, Lynnerup, N., et al. (2003) Distribution patterns of postmortem damage in human mitochondrial DNA. *Am J Hum Genet* 72, 32–47.

56. Irwin, J.A., Saunier, J.L., Niederstätter, H., Strouss, K.M., Sturk, K.A., Diegoli,T.M., 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.

57. Den Tex, R.J., Maldonado, J.E., Thorington, R., and Leonard, J.A. (2010) Nuclear copies of mitochondrial genes: another problem for ancient DNA. *Genetica* 138, 979–984.

58. Binladen, J., Gilbert, M.T., Campos, P.F., Willerslev, E. (2007) 5′-tailed sequencing primers improve sequencing quality of PC products. *Biotechniques* 2, 174–176.

59. Gabriel, M.N., Huffine, E.F., Ryan, J.H., Holland, M.M., and Parsons, T.J. (2001) Improved MtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy. *J Forensic Sci* 46, 247–253.

Chapter 21

Interpretation Guidelines of mtDNA Control Region Sequence Electropherograms in Forensic Genetics

Manuel Crespillo Marquez

Abstract

Forensic mitochondrial DNA (mtDNA) analysis is a complementary technique to forensic nuclear DNA (nDNA) and trace evidence analysis. Its use has been accepted by the vast majority of courts of law around the world. However for the forensic community it is crucial to employ standardized methods and procedures to guaranty the quality of the results obtained in court. In this chapter, we describe the most important aspects regarding the interpretation and assessment of mtDNA analysis, and offer a simple guide which places particular emphasis on those aspects that can impact the final interpretation of the results. These include the criteria for authenticating a sequence excluding the contaminant origin, defining the quality of a sequence, editing procedure, alignment criteria for searching the databases, and the statistical evaluation of matches. It is not easy to establish a single guide to interpretation for mtDNA analysis; however, it is important to understand all variables that may in some way affect the final conclusion in the context of a forensic case. As a general rule, laboratories should be cautious before issuing the final conclusion of an mtDNA analysis, and consider any significant limitations regarding current understanding of specific aspects of the mtDNA molecule.

Key words: Mitochondrial DNA, Interpretation, Forensic, Database, Heteroplasmy

1. Introduction

Analysis of mtDNA (regions HVS-I, HVS-II, and eventually HVS-III) for forensic purposes is currently routine practice in forensic laboratories, especially when the amount and/or quality of DNA obtained is insufficient to conduct conventional analysis of short tandem repeats (STRs). The interpretation and evaluation of the results of mitochondrial DNA is the final step in the analysis. This is not a trivial issue. The analysis of mtDNA and the other methods carried out in laboratories must have been internally validated according to ISO/IEC 17025, as a basic step toward ensuring the

results issued. It has been shown that applying good laboratory practice (GLP) helps to minimize potential errors during analysis. However, other factors, such as the phenomenon of heteroplasmy (position and length), editing of the sequence, sequence quality requirements, databases and database searches, the statistical treatment of matches, the higher mutation rate of mtDNA compared with nuclear DNA (nDNA), and the current lack of knowledge of nucleotide mutation rates—mean that the chance of an mtDNA result being incorrectly interpreted is more likely than under autosomal STR marker analysis, with the consequent legal impact. In some cases, ambiguities in interpreting mtDNA may mean that two analysts working on the same case reach different conclusions; this is because, in certain situations, there is a fine line between assigning the same origin to two samples and excluding that same origin. It is therefore crucial to understand the different aspects that will lead a forensic expert to issue one conclusion over another in this type of analysis.

The international community has attempted to address this problem through various guidelines and recommendations (1–3). However, despite this, this interpretation stage is still the most complex within the entire mtDNA analysis.

Finally, it is true that forensic laboratories have today achieved a high technical standardization and they have a wide range and variety of computer tools that enable them to reach the levels of reliability that this type of analysis, by its nature, requires. In parallel, there have been significant advances in phylogenetic knowledge of mtDNA while there has also been growth in the quantity and quality of population databases. However, despite all these, there is currently no software available that is capable of performing robust and automatic evaluation of mtDNA sequencing data without the manual intervention of an analyst. Adequate training and experience of the analyst therefore continues to be essential in order to handle this type of study successfully.

2. Methods

Interpreting the results of an mtDNA analysis aims to reach conclusions, within the context of a forensic case, regarding whether, given one known sample and one that is questioned, they could come from the same person or maternal origin. However, before moving to the results interpretation phase itself of an mtDNA analysis in a forensic case, two prior compulsory steps must be carried out. These aim to assess the sequence generated and attempt to answer two questions: first, whether or not the sequence is the product of contamination; and second, whether or not the sequence is of sufficient quality to be considered acceptable according to our

laboratory's quality standards. It is during the sequence editing phase (point 3) that such matters are ultimately resolved. Therefore, we must:

1. Determine the authenticity of the sequence in order to rule out contamination.
2. Establish the sequence quality.

2.1. Quality Assurance

Avoiding contamination during genetic analysis is a priority for forensic laboratories; even more so when dealing with mtDNA analysis, justified by the molecule characteristics and the analytical methodology used. Therefore international recommendations regarding facilities and analytical procedure designed to minimize contamination during the process (1) must be adhered to. These recommendations include the use of gloves, gowns, and masks when handling samples as well as when conducting tests in laminar flow or dead space hoods, aerosol-resistant pipette tips, preamplification areas are physically separated from postamplification areas, the use of reagent blanks and negative controls.

The analysis of negative controls—both extraction reagent blank and negative PCR control—is the first step in interpreting the mtDNA sequence. Determining the amount of amplified material through analysis of capillary electrophoresis (2100 Bioanalyzer®. Agilent Technologies) (see Notes 1 and 2) will enable an initial decision to be taken in the process of mtDNA results interpretation. Laboratories performing mtDNA analysis must establish thresholds that limit procedures from continuing when DNA in the reagent blanks and/or negative controls is detected. Some authors consider that if the amount of quantified DNA for the blanks exceeds 10% (4, 5) of the value obtained in the test samples, then the process should be stopped and restarted at the step where the contamination was introduced. If, however, reagent and amplification blanks are estimated below 10%, then the sequencing process can proceed. One study has demonstrated that the 10:1 rule is conservative and reliable (6); therefore, low levels of contamination can be tolerated. However, the final assessment of these negative controls should be done after sequencing them during the editing phase and in the light of the sequencing results obtained for the coextracted and/or amplified test samples along with the these checks.

2.2. Establishing the Quality of the Sequence

The vast majority of forensic laboratories today analyze mtDNA through sequencing, using dye terminator sequencing chemistries (BigDye Terminator® v 3.1 Applied Biosystems) and subsequent detection through capillary electrophoresis. A number of circumstances—depending on the sequencing reaction, electrophoresis conditions (e.g., capillary, buffer, and polymer), analysis software or the DNA sequence itself—can generate poor quality sequences.

Various artifacts such as background noise, low peaks, messy data at the beginning of the sequence or throughout the sequence, irregular spacing, position or length heteroplasmy, dye blobs, spikes can compromise the quality of the sequences studied. The analyst must know this problem and have the tools to correct them. Described below are some of the most common problems and factors along with their origin and potential solution (7).

1. Failed reaction

For many failed reactions analyzed data is not present because the signal strength is below the threshold for analysis. There can be a number of causes to explain this situation—including the small amount of template, poorly designed or inadequate concentration of primers, faults in the PCR reaction, or loss of extension products during cleanup reaction, as well as problems during the injection step. These circumstances must be evaluated in order to correct this type of sequence.

2. Background noise

The most common cause comes from the use of an insufficient amount of template DNA during sequencing reaction, although other causes such as the simultaneous use of multiple primers or type of DNA in the same tube can generate this type of sequence, as well as DNA template that contains impurities, such as salts or ethanol. On the other hand, we must not forget that poor electrophoresis (buffer, polymer), the use of parameters or improper use of a matrix can also cause this type of problem. The result is the generation of multicolored peaks in certain positions, and the correct allocation base by the software is not possible, with these positions normally being characterized as "N" (see Fig. 1). The degree of background noise, the analyst's expertise, and the internals of the laboratory protocols will determine the interpretive value of these sequences.

3. Dye blobs

A fairly common factor when working with dye terminator sequencing chemistries is the appearance in the first sequence section (approximately up to 100–140 bp) of dye blobs as a result of traces of unincorporated terminators that have not been elimi-
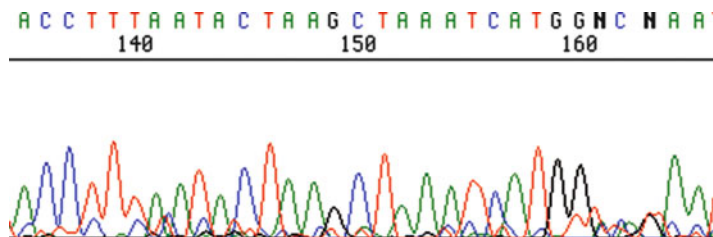
Fig. 1. Electropherogram with background noise.
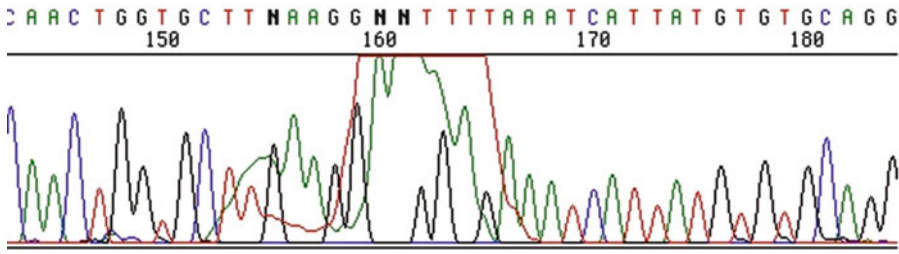
Fig. 2. Traces of unincorporated terminators that have not been removed can cause dye blobs.
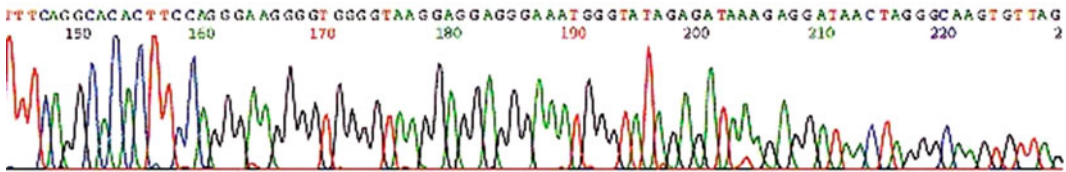

Fig. 3. Note the gradual loss in the intensity of the signal of the peaks.

nated during the purification step. Dye blobs generally occur over the top of the sample peaks without distorting the true DNA sequence (see Fig. 2). Laboratories should use efficient purification systems to eliminate the remaining terminators.

4. Gradual loss of signal

The gradual loss of signal sometimes prevents complete editing of the sequence. This may be due to a number of causes, but the most likely appears to be an imbalance between the amount of DNA and the concentration of primers used for the PCR reaction, an imbalance of any of these components can cause this type of artifact. It is therefore very important to adjust properly the amount of template DNA to be sequenced and the concentration of primer. Moreover, although less likely, excessive salt can also cause this type of sequencing patterns since the salts have an inhibitory effect on Taq polymerase, which can translate into an overabundance of short fragments at the expense of longer fragments (see Fig. 3).

5. Multicolor narrow peaks anywhere in the electropherogram

The appearance of condensed and narrow peaks where different colors of the fluorochromes coexist, and which often mask one or two real nucleotides, may occur anywhere in the electropherogram, and often significantly exceed the height of the surrounding peaks. In these positions, the software assigns values "N" (see Fig. 4). This type of artifact is usually associated with the presence of small bubbles within the polymer matrix, or even small residues of dried polymer suspended in the polymer. It is therefore necessary to work with fresh polymer and view the presence of such
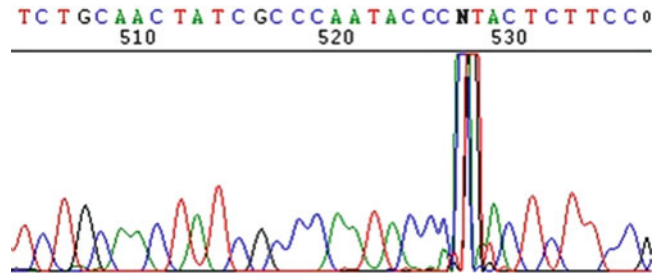
Fig. 4. Multicolor narrow peak.
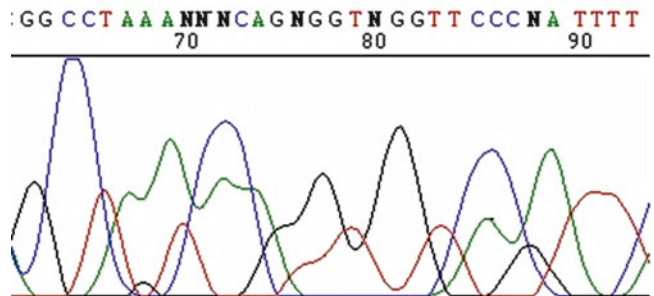
Fig. 5. Broad peaks.

microbubbles prior to electrophoresis. In any event, and when editing the sequence, there should be a new electrophoresis.

6. The appearance of broad peaks with the consequent loss of resolution

Broad and imprecise bands appear in the electropherogram, often covering the space corresponding to several bases. The software incorrectly assigns short homopolymeric tracts (2 or 3 nucleotides—for example TT, CCC…) or "Ns" (see Fig. 5) to the space corresponding to the band. The cause is usually due to the presence of a bubble in the capillary or capillary blockage by some kind of artifact, and the presence of excess salts in the injected sample. Often a reinjection of the sample or cleansing it thoroughly solves the problem.

7. Peaks widened at the base

A strong characteristic of overused capillaries, usually above 150 punctures. At this point the manufacturer's recommendation is to change the capillary (see Fig. 6).

8. Using the analysis software

If the data are analyzed with the wrong software settings (use of a wrong mobility file, incorrect or poor-quality instrument matrix file), the resulting electropherograms will show overlapping peaks and gaps between peaks rather than the evenly spaced peaks characteristic of correctly analyzed data and this causes misassigned

GCCACC A NGGA ACNCA GGTNTNACATANGANATGACNGCCNG
         110   120   130   140
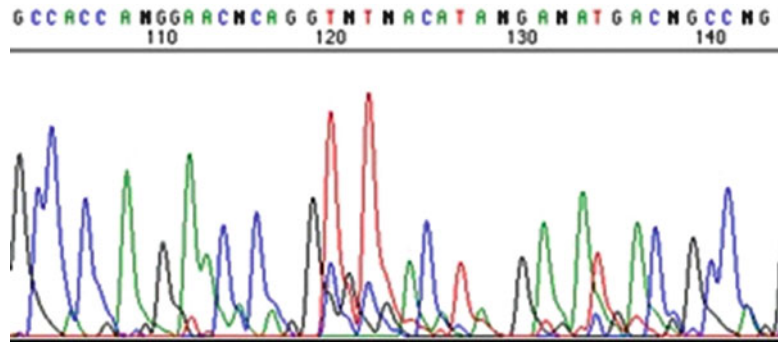
Fig. 6. Overuse of the capillary can cause peak widened at the base.

nucleotides. The result is poor quality electropherograms; for example, the use of an incorrect mobility module is characterized by the presence of abundant background noise, the appearance of overlap in certain bases and the loss of some other assignment. Similarly, the use of a poor quality or incorrect matrix also tends to cause a degree of background noise to appear that affects the quality of the sequence. The result is misassigned bases or "N." As recommended in these situations, the parameters used in the analysis should be checked.

9. Poli C-stretch

The homopolymeric C-stretch region located at nucleotide positions 16184–16193 in HVS-I shows a T (16189) placed between two tracts of Cs. Length heteroplasmy likely results from slippage replication after a T to C transition occurs in position 16189. The impact of the mixture of the length variants generates that the sequence quality drops after the string of cytosine residues.

A similar situation occurs in HVS-II C-stretch when insertion of cytosines occurs in the tract 303–310 or a transition T to C occurs at position 310, generating a, as "out of phase sequence carryover" downstream of the homopolymeric tract, which prevents a double complete reading of the fragment analyzed. To solve the situation different primer combinations are used in order to recover sequence information from both sides of the homopolymeric stretch of cytosines or alternatively the same strand may be sequenced twice in separate reactions to obtain a double lecture of each nucleotide (see Fig. 7).

2.3. Editing Data

Following the amplification reaction and mtDNA sequencing, the capillary electrophoresis and after the raw sequencing data were analyzed (DNA Sequencing Analysis Software®, Applied Biosystems), data is dumped to the assembly software and sequence comparison.

During this editing stage, we can take decisions about the suitability of the sequences to be subjected to comparative study
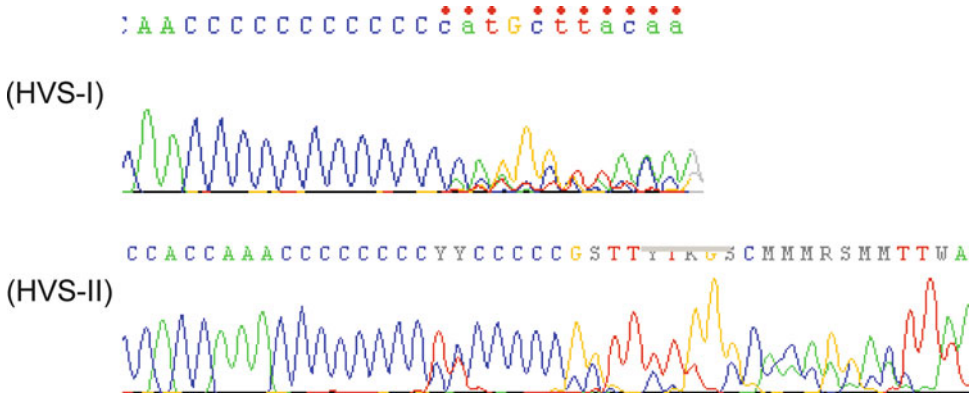
Fig. 7. PoliCs stretch appears in both regions HVS-I and HVS-II producing an abrupt crash in the quality of sequence.

and subsequent assessment or, conversely, if they require further investigation. Also, by editing, we will evaluate definitively if the reagent blanks and amplification controls can be treated as negative (see Fig. 8).

We consider it important that this editing step be carried out by two independent analysts; this will reduce the possible arbitrariness in the allocation of certain bases that sometimes occurs. A sequence analysis is considered correct when both analysts arrive at the same conclusion. If not, additional tests are needed to confirm.

Various programs are now available that allow the assembly, alignment, and comparison of the sequences. The usefulness of such programs in allocating the bases is without question, but unfortunately these programs do not always allow for fully automatic and error-free editing. It is therefore up to the analyst to complete the final editing of the sequences and evaluate whether they attain sufficient reliability and quality to be subjected to a comparative study and subsequent interpretation.

1. The raw data are imported into a comparison program (Sequencher™-Gene Codes-, SeqScape®-Applied Biosystems-). The majority of programs currently used allow both strands (heavy and light) to be aligned and the subsequent generation of a consensus sequence, which can be compared with other sequences as well as with the revised Cambridge reference sequence (rCRS) (8).

2. At this point, the result of the sequence generated from the reagent blank and negative amplification is assessed. If the result of the sequencing enables an interpretable result to be generated, in view of the results, the source of contamination should be assessed and the process restarted at the stage where it is suspected that the contamination occurred. If on the other hand the result of sequencing blanks (reagents and amplification) is negative, we are able to ensure the absence of contaminants in the process, and therefore the authenticity of the sequences of the samples.
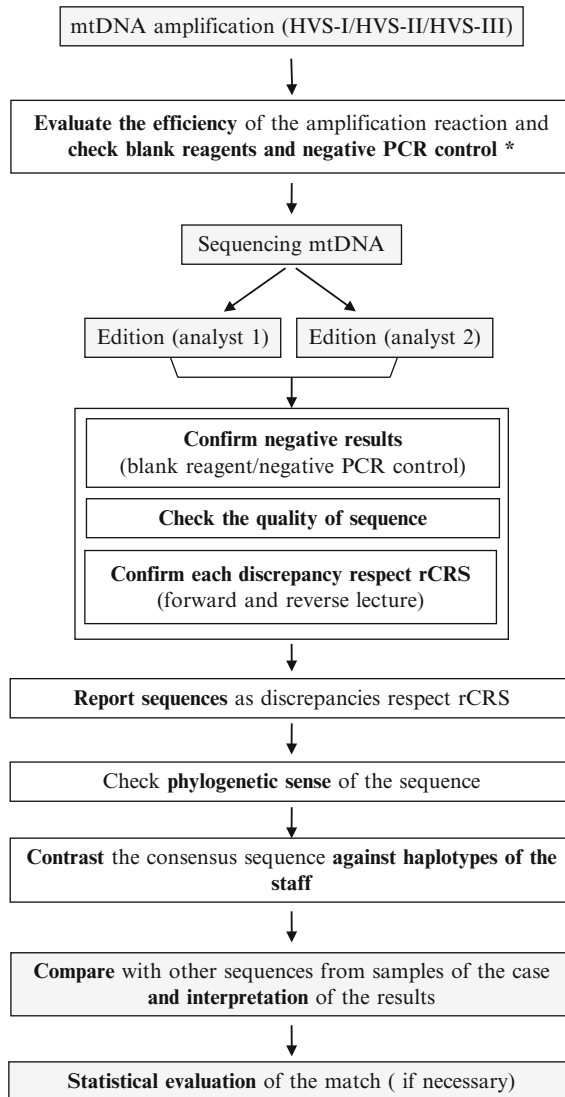
mtDNA amplification (HVS-I/HVS-II/HVS-III)

↓

Evaluate the efficiency of the amplification reaction and
check blank reagents and negative PCR control *

↓

Sequencing mtDNA

Edition (analyst 1)   Edition (analyst 2)

Confirm negative results
(blank reagent/negative PCR control)

Check the quality of sequence

Confirm each discrepancy respect rCRS
(forward and reverse lecture)

↓

Report sequences as discrepancies respect rCRS

↓

Check **phylogenetic sense** of the sequence

↓

Contrast the consensus sequence **against haplotypes of the
staff**

↓

Compare with other sequences from samples of the case
and interpretation of the results

↓

Statistical evaluation of the match (if necessary)

Fig. 8. Steps involved in performing mitochondrial DNA analysis. In general, after each control step, and depending on the result, the laboratory must take decisions based on internal and validated criteria. *Acceptance of the results of negative controls will depend on the method used for analysis of PCR products (capillary electrophoresis or agarose gel) as well as the acceptance criteria established by the laboratory.

3. General supervision of the sequence generated in the samples, evaluating the position "N" that can occasionally occur, and editing any that are permitted as described (see point 2). The presence of an abundance of "N" will not offer proper editing, and in this case will require a determination of the cause and try to correct the problem (see Note 3).

4. rCRS discrepancies assigned by the software should be reviewed manually one by one by the operator and require confirmation of the base allocated on both strands by well-defined peaks and

zero or negligible background noise levels. Reviewing and correcting of any erroneous discrepancies that the program can create "artificially" because of the poor quality of the sequence (see Note 4).

5. Assessment and confirmation of possible position heteroplasmic: Sometimes the software can assign a value of "N," for example when the two populations of mtDNA molecules appear approximately in proportion 1:1 in a particular position, or at other times the base assigned for the majority population of mtDNA molecules present at that position. In the latter case we must be careful because if the majority population does not differ from rCRS, heteroplasmic position may go unnoticed. The assignment of a position heteroplasmy must be confirmed in both strands (see Notes 5 and 6).

6. Length heteroplasmy: The poliCs tracts present in both HVS-I and HVS-II, as described above, are prone to length heteroplasmy in these fragments. Such situations involve achieving a double reading of the strands either through the use of internal primers or alternatively with a double sequencing reaction for the same strand. When it is not possible to determine accurately the exact number of cytosine residues present, no attempt should be made to count the number of residues (for interpretation purposes), and all comparisons will treat this region as having the same number of cytosines.

7. The procedure described in the preceding paragraphs (2.3.1–2.3.6) should be repeated independently by another specialist, thereby obtaining a second edition that will be checked against the first. Where this coincides with the first edition, proceed to next step.

8. At this point, we need to determine the phylogenetic sense of the sequence; in order to discard errors attributable to human and technical issues (see Note 7), if the sequence has that phylogenetic sense then we proceed to next point.

9. As a last step before final acceptance of the sequence, the sequence will be contrasted against the haplotypes of laboratory personnel involved in the analysis. If no match to any member of staff is found, then we can accept the sequence. Otherwise, and depending on the haplotype obtained, the laboratory must take steps to confirm the result, including carrying out a repeated analysis of evidence from the start, provided that the sample permits, for confirmatory purposes.

2.4. Interpretation

After editing and reviewing the sequences resulting from the mtDNA analysis, a comparison between the sequences obtained from a questioned and the known sample is required. The result of comparative analysis can lead to situations such as

1. A full match between the doubted and the reference sample: In this case we cannot exclude the possibility that both samples come from the same person or persons with the same maternal parentage; in this case, we recommend performing a statistical evaluation of the match (discussed below). Even if both samples share some heteroplasmic position, this represents the cooccurrence of an additional unlikely event, and therefore its presence may increase the strength of the evidence (see Note 8) and one cannot exclude two samples as originally coming from the same source or maternal lineage.

2. When there are more than two differences between the sequences from doubted and undoubted samples we can conclude that both samples come from different people or people with different maternal kinship. But to reach this conclusion we must consider these positions as unequivocally different, it is necessary to take into account aspects such as tissue specificity and mutation rates (see Notes 9 and 10).

3. Differences in a single base between the known and questioned sample. In a heteroplasmic position: When a heteroplasmic position that not is present in the known sample is detected in the questioned sample (or vice versa) we cannot exclude the possibility that the doubted and reference samples are from the same person or persons with the same maternal kinship, since both samples share at least one of the two subpopulations of mitochondria molecules. It is therefore advisable to carry out a statistical evaluation of the match. In a length heteroplasmy: if the reference and the doubted sample only differ in length heteroplasmy, that difference alone should not be considered for the purposes of interpretation and therefore in this case we cannot exclude the possibility that the doubtful and reference samples come from the same person or persons with the same maternal parentage; it is advisable to carry out a statistical evaluation of the match. Homoplasmic differences. When both samples, reference and doubted, vary in a position where there is no evidence of heteroplasmy is undoubtedly a complicated situation in the process of interpretation, which could change the final interpretation from a match in the origin of to noncoincidence (see Notes 11 and 12). In this situation, whenever possible at least these steps should be followed.

   Additional known samples (blood, saliva, hair) can be sequenced to determine if the heteroplasmy is visible in other tissues. Obviously, further testing cannot always be performed on a crime scene sample of limited quantity, but it can prove helpful for interpretation of known samples. Special attention must be taken with hairs (see Note 13) and it is advisable to analyze different portions of the same hair.

2.5. Statistics

When the sequences of the reference and doubted samples match, and therefore cannot be excluded as coming from the same origin, it is desirable to convey some information about the weight of the evidence. For this purpose, the laboratory can employ several options:

1. Estimate the rarity of the sequence by expressing the number of times that a particular sequence is found in a database (see Notes 14–17). This method is referred to as the "counting method."

2. By calculating the frequency of that haplotype. Considering that not all haplotypes are represented in a database, sampling error correction needs to be applied. Balding and Nichols (9) propose that adding the haplotype of the doubted and reference sample to the database the sampling error can be roughly corrected.

$$p = \frac{x+2}{n+2} \tag{1}$$

where x is the number of observations of this haplotype in a database of size n.

Holland and Parsons (10) proposed the use of a confidence limit of 95%.

$$p = 1 - \alpha^{1/n} \tag{2}$$

where α is the value of 0.05 for 95% confidence interval and n is the size of the database.

When one or more identical sequences are found in the database, it is advisable to use confidence intervals using the following formula (10).

$$p \pm 1.96\sqrt{\frac{p\,(1-p)}{n}} \tag{3}$$

Calculating p as

$$p = \frac{x}{N}$$

where x is the number of times a profile has been observed in a population and N is the number of profiles in that population.

As recorded by Tully et al. (2), we conclude that, where the sample has not been observed in a database for the calculation of the probability of that haplotype in question, then use either the formula proposed by Balding and Nichols (1) or that of Holland and Parsons (2). However, where this haplotype appears in the database one or more times, the Balding and Nichols formula can be used or alternatively by confidence intervals (3).

3. However, it is recommended that the frequency of a particular haplotype is expressed by its LR (likelihood ratio), and will generally be proposed as follows:

$$LR = \frac{\text{Probability of obtaining the SSD and SSR/DS and RS come from the same maternal line}}{\text{Probability of obtaining the SSD and SSR/DS and RS do not come from the same maternal line}}$$

SSD: Sequence of the doubted sample

SSR: Sequence of the reference sample

DS: Doubted sample

RS: Reference sample

If the questioned and reference sequences match exactly at all points of disagreement with respect to rCRS, the results may be expressed or using the "counting method" or better still expressed in the form of LR, so that the numerator of the above formula would be 1 (since we raised this hypothesis under the assumption that both the questioned sample and the reference sample come from the same person or persons with identical maternal lineage) then in the denominator we would have the p obtained by using sampling corrections proposed by Balding and Nichols or those proposed by Holland and Parsons or statistical approach of EMPOP.

3. Notes

1. This technique compares the amount of DNA in the PCR product to a known DNA standard to determine the concentration of the DNA in the PCR-amplified sample. Negative controls (reagents and amplification) are also included to indicate whether any exogenous DNA is present. Furthermore, detecting amplicons using capillary electrophoresis (CE) has three major advantages over quantification using agarose gel: (a) high sensitivity, (b) high precision, and (c) ability to detect the presence of length heteroplasmy. This last matter is a significant saving of time as it is already known a priori that samples will require additional analysis to resolve the confusing readings that are usually generated when length heteroplasmies are present.

2. As an alternative to quantifying amplified products through capillary electrophoresis, some laboratories compare amplification results with a standard length and known quantity, using a submarine electrophoresis analysis in agarose gel. This technique offers a lower resolution, and therefore being less reliable

decision making in reference to reagent blanks and amplification negatives. In this case it is helpful to analyze up to the end point (sequencing) the reagent and amplification blanks and take decisions at the end of the process, with the consequent increase in time.

3. If possible, the sequencing comparison software must be configured in accordance with the quality requirements set by the laboratory based on the validation of the method performed (e.g., defining the parameters for identification of mixed bases, clear range or filter settings).

4. By convention, the nucleotide positions of the human mtDNA genome are numbered from 1 to 16569 according to the rCRS. Likewise, for reporting purposes, the sequences are described as discrepancies with respect to the sequence of Anderson revised (rCRS). So when it detects a difference between the analyzed sequence and rCRS the nucleotide position is cited followed by de base present at that site. Insertions are designated by a decimal point and a "1" ("2" if a second insertion is reported, and so on) after the lower numbered nucleotide of the two between which it has inserted. However if an insertion occurs within a homopolymeric tract (poliCs) the exact location of the insertion is unknown, then the standard assume that the insertion has occurred at the highest numbered end of the stretch. Deletions are noted by listing the position deleted followed by a "d." At confirmed heteroplasmic positions, IUPAC codes can be applied. If the position has not yet been confirmed as heteroplasmic, then it can be reported as "N."

5. Some authors propose (2) that in order for a mutation to be detected by sequencing, it must be present at a level approaching 20% to be distinguished from background. However the chance of detection of heteroplasmy is dependent upon the sequencing chemistry used. In addition, detection may be more efficient at certain nucleotide positions than at others, and differences in detection may be also observed between both strands of mtDNA. Sometimes heteroplasmy detected in one strand can not be confirmed in the complementary and sometimes even in a second reaction. In these cases, the recommendation states that the position would be designated as ambiguous (N) in the relevant extract.

6. It is necessary to distinguish between a sequence in the presence of heteroplasmy, or a mixture of sequences. Individuals with two heteroplasmic positions have been reported (11); however, heteroplasmy at three positions is certainly quite rare and this situation should lead us to think that the sequence obtained is the product of a mixture, either by contamination or from the original source material. Interpreting mixtures in forensic cases is not currently common practice by the laboratories.

7. Several publications have shown the propensity for human error in mtDNA analysis (12, 13) as mix-up can be generated when handling the tubes or in the course of generating samples sheets prior to the electrophoresis step, misinterpretation of sequence raw data, "phantom mutations" caused by artificial signals in the sequencing electropherograms or clerical errors during data transcription. Phylogenetic evaluation of sequences helps to uncover these kinds of errors. Phylogenetic analysis of the sequence obtained has been demonstrated to be an effective tool that enables this type of error to be detected (14, 15). However, the assignment of haplogroups to mtDNA profiles is difficult, especially if only control region information is available, and most haplogroups are defined by coding region polymorphisms. Currently there are bioinformatic tools that automatically allow to forensic community to estimate the most probable mtDNA haplogroups of the mtDNA control region sequences (DNAmanager) (16). A powerful web application (Haplogrep) (17) has recently been developed based on Phylotree (18) that enables accurate estimation of the haplogroup.

8. The current knowledge regarding the mutation rate at each position in mtDNA molecule is still limited. Therefore, the lack of current information concerning knowledge about mutation rates at each position prevents statistical processing from being carried out of matches between samples with common heteroplasmic positions.

9. Various studies indicate that segregation of mutations occurs at different rates in different tissues (e.g., highest in muscle or hair) (19). This phenomenon is due to the different mechanisms by which cells are generated in different tissues. Therefore the sources of the tissues investigated should be taken into consideration, because differences in mtDNA sequence due to mutations seems to be more likely for example between hair and blood than between two blood samples taken from the same individual.

10. Positions that are more prone to mutation than others have been reported (19–21), and these points have been defined as "hotspots," including the following positions in HVR-I: 16093, 16129, 16153, 16189, 16192, 16293, 16337, and 16309 and 72, 153, 189, 207, and 279 in HVR-II. Therefore, differences at phylogenetically stable positions provide more evidence for exclusion than differences at known hotspots.

11. It is well established that the mutation rate in mtDNA is significantly higher than that which occurs in nDNA. Mutations are passed between generations in varying ratios and segregate during development and later life; mutations also accumulate and segregate during the lifetime of an individual. This means that mixtures of two or more subpopulation molecules of

mtDNA can occur within individuals. Then, we can observe discrete differences (1 or 2 positions) between different tissues of the same individual. This situation also can be shown between individuals related through the maternal line (maternal relatives). These aspects are reviewed extensively by Holland and Parsons (10).

12. The bottleneck theory (22) may explain the appearance of different tissues in an individual that do not match within the mtDNA sequence, occasionally in a homoplasmic position. Again, it is necessary to assess the divergent base and determine whether it is a "hotspot" position. It is also necessary to assess the type of tissue involved in the analysis and occurrence of this haplotype in the database. Logically if the degree of variability between the two samples occurs within a stable nucleotide, then we will be inclined to think that the doubted and undoubted samples have different origins. However, if the difference appears in a "hotspot" base (e.g., 16093), then we will tend to think to both samples—doubted and undoubted—could come from the same maternal origin. Again, current knowledge on mutation rates at each position prevents a more accurate assessment of such situations.

13. Several authors have highlighted the variability of sequences that can be found within a hair (23, 24); the hair's own histogenesis could explain this fact (25). Therefore, special caution in interpreting the results of this type of samples is advisable.

14. To avoid errors of transcription, it is desirable that the laboratory has systems that enable searches on population databases via automatic download of the query sequence, and therefore without manual intervention by the analyst. There are now programs that enable data to be imported from alignment and analysis software through different formats (26).

15. Numerous publications have highlighted errors in the data contained in databases (12–14), essentially the error rates are concentrated in phantom mutations, base misscoring, base shifts, and artificial recombination. Having high quality database is crucial in order to make a reliable estimate of the frequency for a random match. There are currently multiple databases of mtDNA (27, 28) for consultation of the scientific community, but in recent years the European DNA profiling group (EDNAP) has been launched to construct a mtDNA database of reliable population data. Currently, EDNAP mitochondrial DNA Population Database (EMPOP) has 14,847 haplotypes and today this database is one of the most used by the forensic science community.

16. Database searches are usually done by alignment with respect to the rCRS. This task is straightforward for the vast majority
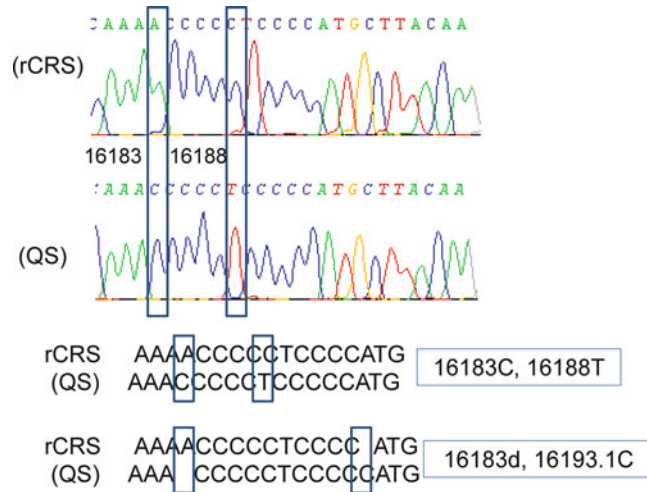
Fig. 9. Different alignment for a sample (*QS* questioned sample) with respect to the rCRS.

of nucleotide positions, but is difficult in regions displaying length variation. Different alignments of the same sample can result in mismatches on several nucleotide positions (see Fig. 9). The choice of one or other option can have a significant impact on the final estimates of the frequency of the haplotype. Wilson et al. (29, 30) in 2002 proposed rules for most parsimonious sequence alignment with respect to the rCRS; however, Bandelt and Parson (31), in view of the significant increase on the knowledge of the phylogenetic distribution of mtDNA propose an alternative approach to the alignment.

17. Phylogenetic alignment in EMPOP database suggests three (main) rules (31):

- Phylogenetic rule: Sequences should be aligned with regard to the current knowledge of the phylogeny. In the case of multiple equally plausible solutions, one should strive for maximum (weighted) parsimony. Variants flanking long C tracts, however, are subject to extra conventions in view of extensive length heteroplasmy.

- C-tract conventions: The long C tracts of HVS-I and HVS-II should always be scored with 16189C and 310C, respectively, so that phylogenetically subsequent interruptions by novel C to T changes are encoded by the corresponding transition. Length variation of the short A tract preceding 16184 should be notated in terms of transversions.

- Indel scoring: Indels should be placed 3′ with respect to the light strand, unless the phylogeny suggests otherwise.

Acknowledgments

References

1. Carracedo A, Bär W, Lincoln P et al (2000) DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int 110:79–85.

2. Tully G, Bär W, Brinkmann B 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.

3. Scientific Working Group on DNA Analysis Methods (SWGDAM) (2003) Guidelines for mitochondrial DNA (mtDNA) nucleotide sequence interpretation. Forensic Sci Commun 5 (2). http://www.fbi.gov/hq/lab/fsc/backissu/april2003/swgdammitodna.htm.

4. Wilson MR, DiZinno JA, Polanskey D et al (1995a) Validation of mitochondrial DNA sequencing for forensic casework analysis. Int J Legal Med 108:68–74.

5. Wilson MR, Polanskey D, Butler J. et al (1995b) Extraction, PCR amplification, and sequencing of mitochondrial DNA from human hair shafts. Biotechniques 18:662–669.

6. Stewart JE, Aagaard PJ, Pokorak EG et al (2003) Evaluation of a multicapillary electrophoresis instrument for mitochondrial DNA typing. J Forensic Sci 48:571–80.

7. Automated DNA Sequencing Chemistry Guide. Applied Biosystems. https://products.appliedbiosystems.com.

8. Andrews RM, Kubacka I, Chinnery PF et al (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147.

9. Balding D, Nichols R (1994) DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 64: 125–140.

10. Holland M, Parsons T (1999) Mitochondrial DNA sequence analysis-validation and use for forensic casework. Forensic Sci Rev 11:21–50.

11. Tully LA, Parsons TJ, Steighner RJ et al (2000) A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region I of the human mtDNA control region. Am J Hum Genet 67: 432–443.

12. Bandelt H-J, Quintana-Murci L, Salas A et al (2002) The fingerprint of phantom mutations in mtDNA data. Am J Hum Genet 71: 1150–1160.

13. Bandelt H–J, Lahermo P, Richards M et al (2001) Detecting errors in mtDNA data by phylogenetic analysis. Int J Legal Med 115: 64–69.

14. Bandelt H-J, Salas A, Lutz-Bonengel S (2004) Artificial recombination in forensic mtDNA population databases. Int J Legal Med 118: 267–273

15. Bandelt H-J, Kivisild T, Parik J et al (2006) Lab-Specific Mutation Processes. In: Bandelt H-J, Macaulay V, Richards M, editors. Human mitochondrial DNA and the evolution of Homo sapiens. Springer-Verlag, Berlin-Heidelberg. pp. 117–146.

16. Lee HY, Song I, Ha E et al (2008) mtDNA-manager: a Web-based tool for the management and quality analysis of mitochondrial DNA control-region sequences. BMC Bioinformatics 9:473.

17. Kloss-Brandstätter A, Pacher D, Schönherr S et al (2011) HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum Mutat 32: 25–32.

18. Van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30:E386–E394.

19. Calloway CD, Reynolds RL, Herrin GL et al (2000) The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. Am J Hum Genet 66:1384–1397.

20. Mark Stoneking (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet 67:1029–1032.

21. Brandstätter A, Parson W (2003) Mitochondrial DNA heteroplasmy or artefacts – a matter of the amplification strategy? Int J Legal Med 117:180–184.

22. Bendall KE, Macaulay VA, Baker JR, Sykes BC (1996) Heteroplasmic point mutations in the human mtDNA control region. Am J Hum Genet 59:1276–1287.

23. Sullivan KM, Alliston-Greiner R, Archampong FIA et al (1997) A single difference in mtDNA control region sequence observed between hair shaft and reference samples from a single donor. Proceedings from the 17th International Symposium on Human Identification. Promega Corporation, Madison, Wis, 126–129.

24. Wilson MR, Polanskey D, Replogle J et al (1997) A family exhibiting heteroplasmy in the human mitochondrial DNA control region reveals both somatic mosaicism and pronounced segregation of mitotypes. HumGenet 100:167–171.

25. Linch CA, Whiting DA, Holland MM (2001) Human hair histogenesis for the mitochondrial DNA forensic scientist. J Forensic Sci 46:844–853.

26. Weißensteiner H, Schönherr S, Specht G et al (2010) eCOMPAGT integrates mtDNA: import, validation and export of mitochondrial DNA profiles for population genetics, tumour dynamics and genotype-phenotype association studies. BMC Bioinformatics 11:122.

27. Monson KL, Miller KWP, Wilson MR, DiZinno JA, Budowle B (2002) The mtDNA population database: an integrated software and database resource for forensic comparison. Forensic Sci Commu. 4 (2). Available: http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm.

28. Parson W, Dür A (2007) EMPOP-A forensic mtDNA database. Forensic Sci Int Gent 1:88–92. Available on-line http://www.empop.org.

29. Wilson MR, Allard MW, Monson K et al (2002) Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region. Forensic Sci Int 129:35–42.

30. Wilson MR, Allard MW, Monson K et al (2002) Further discussion of the consistent treatment of length variants in the human mitochondrial DNA control region. Forensic Sci Commun. 4:1–8.

31. Bandelt H-J, Parson W (2008) Consistent treatment of length variants in the human mtDNA control region: a reappraisal. Int J Legal Med 122:11–21.

Chapter 22

Capillary Electrophoresis of mtDNA Cytochrome b Gene Sequences for Animal Species Identification

Adrian Linacre

Abstract

The identification of a biological species from an unknown material can be performed using a mitochondrial DNA locus. The cytochrome b (cyt b) gene is one of the most commonly used genetic loci, applied in both taxonomy and forensic science, for the purpose of species identification. The gene is 1,140 basepairs long in most vertebrates and shows little intraspecies variation, but sufficient interspecies variation. The gene is too long for typical PCR-based methods and, as many samples are degraded, only a section of the locus is used. DNA sequences at the start of the gene became the favored section for phylogenetics and species identification. The DNA sequences from any unknown sample from this part of the cyt b gene can be determined and compared to those on a DNA database to determine the most likely species from which the unknown sample originates. The process of DNA amplification, sequencing, sequence comparison, and identification form the basis of this chapter.

Key words: Cytochrome b gene, Mitochondrial DNA, Species identification, Sequence comparison

1. Introduction

DNA typing has been utilized in taxonomy, phylogenetics, and forensic science for the purpose of identification of a species and has recently been reviewed extensively (1–5). The need for species identification in forensic science occurs if there is an allegation that an accused is in possession of a protected species. Legal protection comes from both national legislation and the enforcement of international agreements such as those of the Convention on the International Trade in Endangered Species of Flora and Fauna (otherwise known as CITES). The DNA test has to be able to separate closely related species, yet all members of the same species must share the same DNA sequence. The loci of choice for taxonomy,

phylogenetics, and forensic science are on the mitochondrial genome; being either predominantly the cytochrome b (cyt b) or the cytochrome oxidase I genes (COI). The process of species identification is similar for both genes, but as cyt b has marginally more interspecies variation and lesser intraspecies variation than COI for mammalian species (6), this chapter focuses on the use of the cyt b gene.

The cyt b gene is normally 1,140 bases in length, lies between positions 14,747 and 15,887 on the human mitochondrial genome, and encodes a protein 380 amino acid in length (7–9). The protein is part of the oxidative phosphorylation process in the mitochondria and the functional constraints of the protein result in regions of the protein being highly conserved. This allows for the design of primers in the PCR process that are universal for all mammalian species (7, 8, 10–12), amplifying areas of the cyt b gene that are polymorphic at the sequence level for mammalian species. A number of primer sets have been reported, all of which are designed to amplify the first part of the gene and extend around 400 bases into the coding region. This section of the cyt b gene has been used in the species identification; for a recent discussion on species testing using cyt b, see the paper of Tobe et al. (6). The PCR product is of a suitable size for DNA sequencing. In forensic science, it is important to be confident of the sequence, and confirmation can be achieved by sequencing the DNA in both directions. The confirmed DNA sequence is compared to those registered on a DNA database such as GenBank (http://www.ncbi.nih.gov). As there are over 32 million DNA sequences of cyt b on the database as of June 2010, it is highly likely that the species being examined is already on the database; the only exception being if the species is particularly rare.

2. Materials

1. If using reference material for comparison to an unknown sample, then the sources of this reference DNA need to be from a voucher specimen (see Note 1). The method described below is designed to amplify a 402 bp section of the mammalian mitochondrial genome, giving the advantages high copy number as well as a short fragment, such that degraded DNA may yield an amplification product. This attribute can be important if the reference and questioned samples are of poor quality.

2. PCR primers: the universal primers for PCR amplification are L14724 and H15149 (see Fig. 1), as described by Kocher et al. (8) and Irwin et al. (7). The sequences of primers were designed to correspond to consensus sequences of human, mouse, and cow loci (8), and numbering was according to the human
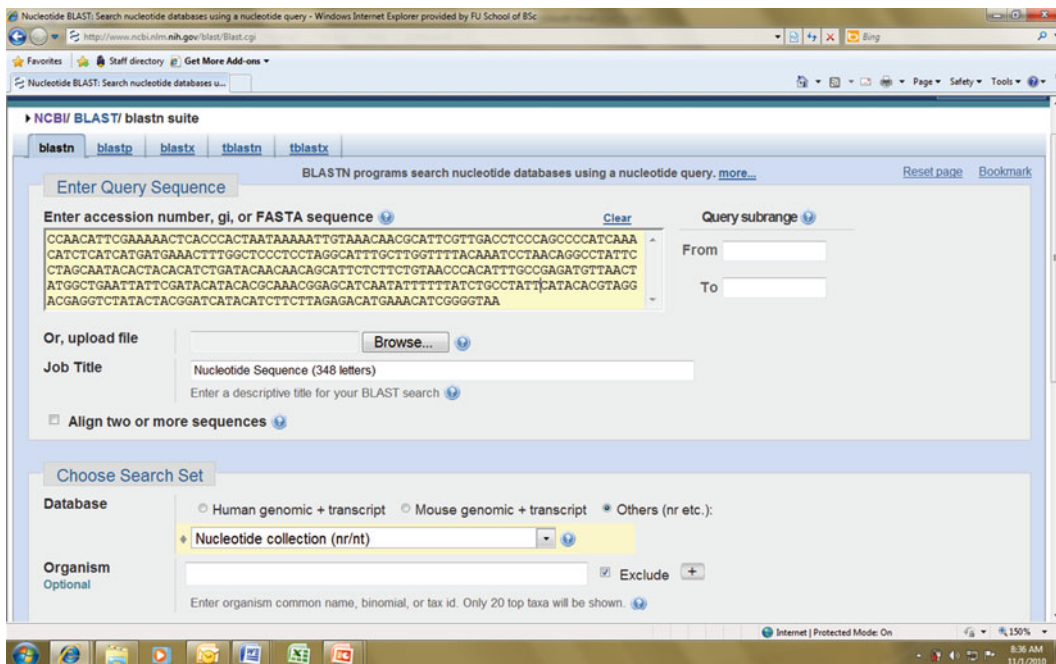
Fig. 1. Screen shot of a similarity search using a 348 fragment of the cytochrome b (cyt b) gene of the Tibetan antelope (*Pantholops hodgsonii*), which is a CITES-protected species.

mtDNA sequences (13). A validation study to show that the section of the cyt b gene amplified by primer sets L14724 and H15149 was universal for mammalian species and that the DNA sequences were species-specific was conducted by Hsieh et al. (10).

3. PCR mix: typically containing reaction buffer ($1\times$ buffer: 10 mM Tris–HCl, pH 8.3, 2.5 mM $MgCl_2$, 50 mM KCl, 0.1% (w/v) gelatin), 200 µM dNTPs, 2.5 unit of VioTaq DNA polymerase, and 0.15 µM each of primers (see Table 1).

4. PCR Machine: any PCR machine will suffice if capable of standard PCR and cycle sequencing.

5. Gel electrophoresis: small gels of less than 10 cm in length are required. A standard power pack and UV transilluminator is required.

6. DNA sequencing: a commercial kit such as the BigDye™ Terminator Kit supplied by Applied Biosystems can be used.

7. Genetic analyzer: equipment such as the Applied Biosystems 310, 3500, or similar CE equipment with fluorescence detection is ideal.

8. Internet access: suitable for navigating to DNA sequence comparison web sites.

Table 1
List of primers used in amplifying a section of the cyt b gene for the purpose of species identification

Location	Sequence (5′–3′)	Size (bp)	References
L14841	AAAAAGCTTCCATCCAACATCTCAGCATGATGAAA	308	(8)
H15149	AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA		
L14724	CGAAGCTTGATATGAAAAACCATCGTTG	425	(7)
H15149	AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA		
L14816	CCATCCAACATCTCAGCATGATGAAA	358	(16)
H15173	CCCCTCAGAATGATATTTGTCCTCA		
L14724	CGAAGCTTGATATGAAAAACCATCGTTG	425	(10)
H15149	AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA		

Note that the position of the primers at the 3′ end is based on the human DNA sequence (13) and will be different for other mammalian species

3. Method

1. PCR Amplification

   PCR amplification can be performed in 50 μL of reaction mixture, which contains 1 ng genomic DNA (see Note 2), 1× reaction buffer, 2.5 unit of DNA polymerase, and 0.15 μM each of primers (see Note 3). Amplification using the 9600 Perkin-Elmer thermal cycler can be conducted using the following conditions: initial denaturation at 95°C for 10 min followed by 35 cycles of 95°C for 45 s, 50°C for 45 s, and 72°C for 90 s (see Notes 4 and 5).

   The success of the PCR can be determined by separating and visualizing the PCR fragments on an agarose gel. If this is performed, it is best to use 5 μL of the PCR, separate on a 2% agarose gel (w/v), and use a size marker appropriate for a DNA fragment of around 400 bp (such as a 1 kb size ladder). Agarose gels are stained typically with Ethidium bromide (add 5 μL of 10 mg/mL to a 50 mL agarose while still molten prior to casting the gel). The approximate yield of PCR product can be determined by the relative brightness of the PCR product compared to the size marker. Remember that this is a sequence polymorphism, so all mammalian species should produce the same size of PCR product and only one PCR product should be produced for each sample.

2. DNA Sequencing

Cycle sequencing of PCR products uses standard methods and is described in (10) and in Linacre and Lee (14). Briefly, the same PCR machine can be used for the sequencing reaction, in this case a 9600 Applied Biosystems thermal cycler, and with the following conditions: 25 cycles of 95°C for 45 s, 50°C for 15 s, and 72°C for 4 min. Sequencing can be performed using the same primers as in the original amplification, L14724 or H15149, using the BigDye™ Terminator Kit (ABI PRISM™ BigDye™ Terminator Cycle Sequencing Ready Reaction Kit). Sequence analysis should be conducted in both directions and any ambiguities determined. In common with the mitochondrial D-loop, heteroplasmy is encountered rarely in the cyt b gene, but is known to occur at other loci (see Note 6).

The PCR products may need to be concentrated and purified prior to separation by capillary electrophoresis. This can either be performed by ethanol precipitation, use of a spin column clean-up kit (such as the MiniElute PCR kit from QIAGEN or the Wizard PCR Clean-Up System from Promega), or a microconcentrator such as Microcon 100.

3. DNA Separation by Capillary Electrophoresis

Any platform capable of detecting the fluorescent dyes and can separate DNA fragments differing in size by one base is suitable. The Applied Biosystems series of capillary electrophoresis platforms range from the single capillary 310 Genetic Analyzer, through the 4 capillary 3130, the 24 capillary 3500 Genetic Analyzer, to 96 capillary 3730xl. Each of these pieces of equipment requires their own settings including run modules and mobility files dye setting. The length of capillary and polymer needs to be suitable for one base resolution as required in DNA sequencing. These pieces of equipment are familiar to many in the forensic science arena as they are used for the separation of STR fragments. It is important to note that the settings used in sequencing and the polymer required are different to those used in the separation of STR fragments.

The fragment reads using the process and the separation platforms described above are typically 600 bases. Given that the fragment in this case is 400 bp, complete coverage is possible in one reaction.

4. DNA Sequence Comparison

The simplest method of identifying the species present is by comparison to a reference DNA sequence database (15, 16). The most commonly used databases are compiled by the NCBI/EMBL/DDBJ consortium and can be accessed at a variety of web sites including http://www.ncbi.nih.gov or http://www.ebi.ac.uk/embl.

The primer sequence should be removed prior to making any search for similar DNA sequences. The sequence can be compared to all the other sequences using the program BLAST (stands for Basic Local Alignment Search Tool). BLAST searches the online database for sequences with similarities to the query sequence and will return a list of the closest matches in order of similarity, with those sequences with the greatest homology listed first.

To use the BLAST tool, first go to the section of the NCBI website page specific for BLAST—http://www.ncbi.nlm.nih.gov/blast/Blast.cgi. Paste the DNA sequence in the section "Enter Query Sequence." Under the heading "Basic BLAST," nucleotide BLAST is selected from the options. This will compare a nucleotide query sequence (i.e., questioned sequence just pasted into the box) with the nucleotide sequences on the database. Make sure that the option below is set to "Nucleotide collection" so that the entire sequence database will be compared to the query sequence. Also, if the sequence is known and it is a commonly occurring species, there is an option on this page to exclude DNA sequences from this particular sequence to the alignment (see Note 7). Figure 1 is a screen shot of such a similarity search using a 348 fragment of the cyt b gene of the Tibetan antelope (*Pantholops hodgsonii*); this species is on Appendix 1 of CITES affording it the highest protection.

Click on the BLAST icon at the bottom right, and within a few seconds, an alignment is provided showing the DNA sequences with the closest similarities.

An example of the BLAST results of the alignment is shown in Fig. 2. In this example, the short sequence has a 100% similarity with the Tibetan antelope (not shown), but the next closest species is the Chinese Blue Sheep (*Pseudosis nayaur*) where there is a 94% similarity over the 348 bases. This 6% dissimilarity is expected if these two DNA sequences come from two different mammalian species (6) where no more that 1.5% dissimilarity was noted between any two samples coming from the same mammalian species using the cyt b gene. There is a high degree of confidence that the unknown sample is from the Tibetan antelope (*P. hodgsonii*) that is a CITES listed and protected species.

4. Notes

1. It is essential that a voucher specimen is used as a reference DNA sequence for comparison to any unknown sample. Voucher specimens can be obtained from collections held by museums or zoological institutes. If voucher specimens are not
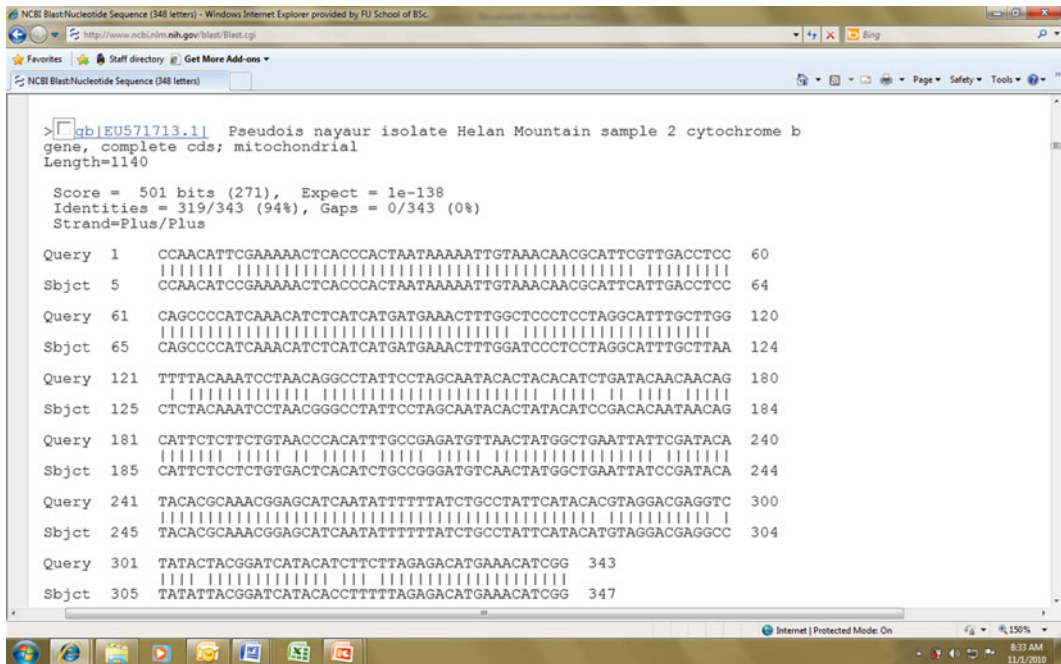
Fig. 2. An example of the Basic Local Alignment Search Tool (BLAST) results.

available, then the next option is to sequence the same locus but from a sample assumed to be of the correct species, in which case it is possible to compare this assumed sequence to the other DNA sequences on the DNA database and it should show a high amount of similarity with other members of the same genus. If the phylogeny of this assumed DNA sequence is as expected, it is a substitute for a voucher specimen. This assumption should be made clear in any subsequent report.

2. Most methods of DNA isolation will isolate whole genomic DNA, comprising both nuclear and mitochondrial DNA. The exact amount of mitochondrial DNA is not known, unless real-time PCR is performed using mitochondrial primers to determine the copy number.

3. The period of time required as the initial denaturation step will alter depending on the DNA polymerase enzyme used. The conditions in the text are for a heat stable enzyme that is activated by the initial high denaturation step, but would be reduced if using a polymerase without this heat activation function.

4. The PCR conditions may alter if using other makes of thermalcycler. The Applied Biosystems 9600 is the validated equipment for PCR of STR multiplex kits provided by this same company and therefore will be found commonly in many forensic science laboratories.

5. 35 Cycles should be sufficient when using 1 ng of isolated DNA. If there is very much less DNA, and hence less mitochondrial DNA, then the number of cycles may be increased.

6. Heterplasmy is the occurrence of two or more DNA sequences at the same single position. Heteroplasmy occurs most commonly in mitochondrial genomes where the cell type is rapidly reproducing and an error is incorporated during replication of the mitochondrial DNA. Hair samples have been shown to exhibit heteroplasmy. The amino acid sequence is highly conserved with most DNA polymorphisms being synonymous mutations and any change in the amino acid sequence is unlikely.

7. There can be hundreds of DNA sequences derived from commonly studies loci such as the cyt b gene for one species. If the aim is to determine if the sequence is not from this well-studied species but from a closely related species, then the option exists to exclude such species from the search.

References

1. E. Alacs, A. Georges, N. FitzSimmons, and J. Robertson, DNA detective: a review of molecular approaches to wildlife forensics, Forensic Science, Medicine, and Pathology, 2009. 3.

2. A. Linacre (2006) Application of mitochondrial DNA technologies in wildlife investigations - species identification, Forensic Science Review, 18(1): p. 1–8.

3. A. Linacre, ed. *Forensic Science in Wildlife Investigations*. International Forensic Science and Investigation Series. 2009, CRC Press: Boca Raton.

4. R. Ogden, N. Dawnay, and R. McEwing, (2009) Wildlife DNA forensics - bridging the gap between conservation genetics and law enforcement, Endangered Species Research.

5. S.S. Tobe and A. Linacre, (2010) DNA typing in wildlife crime: recent developments in species identification, Forensic Science Medicine and Pathology, 6(3): p. 195–206.

6. S.S. Tobe, A.C. Kitchener, and A.M.T. Linacre, (2010) Reconstructing Mammalian Phylogenies: A Detailed Comparison of the Cytochrome *b* and Cytochrome Oxidase Subunit I Mitochondrial Genes, PLoS ONE, 5(11): p. e14156.

7. D. Irwin, T. Kocher, and A. Wilson, (1991) Evolution of the cytochrome b gene of mammals, Journal of Molecular Evolution, 32(2): p. 128–144.

8. T.D. Kocher, W.K. Thomas, A. Meyer, S.V. Edwards, S. Paabo, F.X. Villablanca, and A.C. Wilson, (1989) Dynamics of Mitochondrial-DNA Evolution in Animals - Amplification and Sequencing with Conserved Primers, Proceedings of the National Academy of Sciences of the United States of America, 86(16): p. 6196–6200.

9. R. Kuwayama and T. Ozawa, (2000) Phylogenetic relationships among European red deer, wapiti, and sika deer inferred from mitochondrial DNA sequences, Molecular Phylogenetics and Evolution, 15(1): p. 115–123.

10. H.-M. Hsieh, H.-L. Chiang, L.-C. Tsai, S.-Y. Lai, N.-E. Huang, A. Linacre, and J.C.-I. Lee, (2001) Cytochrome b gene for species identification of the conservation animals, Forensic Science International, 122(1): p. 7–18.

11. S. Pääbo, J.A. Gifford, and A.C. Wilson, (1988) Mitochondrial DNA sequences from a 7000-year old brain, Nucleic Acids Research, 16(20): p. 9775–9787.

12. W. Parson, (2009) Relevance of mtDNA analysis for forensic applications, Rechtsmedizin, 19(3): p. 183–192.

13. R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, and N. Howell, (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nature Genetics, 23(2): p. 147–147.

14. A. Linacre and J.C.-I. Lee (2005), *Species Identification: The Role of the Cytochrome b*

Gene, in *Forensic DNA Typing Protocols*, A. Carracedo, Editor., Humana Press: Totowa, New Jersey, p. 45–53.

15. W. Parson, A. Brandstatter, M. Pircher, M. Steinlechner, and R. Scheithauer, (2004) EMPOP - the EDNAP mtDNA population database concept for a new generation, high-quality mtDNA database, Progress in Forensic Genetics, 10(1261): p. 106–108.

16. W. Parson, K. Pegoraro, H. Niederstatter, M. Foger, and M. Steinlechner, (2000) Species identification by means of the cytochrome b gene, International Journal of Legal Medicine, 114(1): p. 23–28.

Chapter 23

Sequence Analysis of the Canine Mitochondrial DNA Control Region from Shed Hair Samples in Criminal Investigations

C. Berger, B. Berger, and W. Parson

Abstract

In recent years, evidence from domestic dogs has increasingly been analyzed by forensic DNA testing. Especially, canine hairs have proved most suitable and practical due to the high rate of hair transfer occurring between dogs and humans. Starting with the description of a contamination-free sample handling procedure, we give a detailed workflow for sequencing hypervariable segments (HVS) of the mtDNA control region from canine evidence. After the hair material is lysed and the DNA extracted by Phenol/Chloroform, the amplification and sequencing strategy comprises the HVS I and II of the canine control region and is optimized for DNA of medium-to-low quality and quantity. The sequencing procedure is based on the Sanger Big-dye deoxy-terminator method and the separation of the sequencing reaction products is performed on a conventional multicolor fluorescence detection capillary electrophoresis platform. Finally, software-aided base calling and sequence interpretation are addressed exemplarily.

Key words: Canine DNA, mtDNA, Control region, Hairs

1. Introduction

Hairs of nonhuman origin may be among the most frequently collected crime scene traces, but their importance as evidence is often overlooked, as standardized molecular identification methods are still lacking for most species despite great efforts undertaken by the forensic community to establish reliable protocols. In recent years, evidence from domestic dogs has increasingly been used for forensic DNA testing and canine Short Tandem Repeat (STR) analysis has meanwhile become routine practice in dedicated laboratories (1–7). STRs are the markers of choice for traces that contain sufficient amounts of nuclear DNA, but the majority of canine traces secured at crime scenes are naturally shed hairs of telogenic growth

phase and therefore do not contain enough nuclear DNA for successful STR analysis. Similar to human hair mitochondrial DNA (mtDNA), analysis then becomes the most promising method for obtaining molecular information, as mtDNA copy number is also elevated in canine hair.

The approach of forensic canine mtDNA analysis presented here is based on the know-how and experience acquired by human mtDNA analysis (8–11), but is taking the specific features of the dog's mitochondrial genome into consideration. The first complete canine mtDNA sequence has been published by Kim et al. (12) with a total specified length of 16,727 bp (GeneBank accession number: NC_002008, U96639) relative to which all other sequences are reported in difference-based format (13–16). The most polymorphic region of the canine mtDNA is the control region (CR) spanning approximately 1,270 bp in length. It is subdivided into two segments commonly referred to as hypervariable segment 1 (HVS-I) and hypervariable segment 2 (HVS-II). These are separated by a region consisting of 20–30 copies of a 10 bp variable number tandem repeat (VNTR, see Fig. 1).

In this chapter, we exemplify the workflow from stain to mtDNA profile for canine evidence in general and for canine hair shafts in particular starting with careful sample handling and the description of a sample preparation procedure that is minimizing the risk of contamination. In brief, hair material is lysed by adding Dithiothreitol (DTT) and Proteinase K and DNA is obtained by performing Phenol/Chloroform extraction. The amplification and sequencing strategy outlined here can be used for various qualities of mtDNA, but was primarily designed for telogen hair and hair shaft samples as well as degraded blood and tissue samples. For low mtDNA amounts or low quality DNA such as encountered in hair
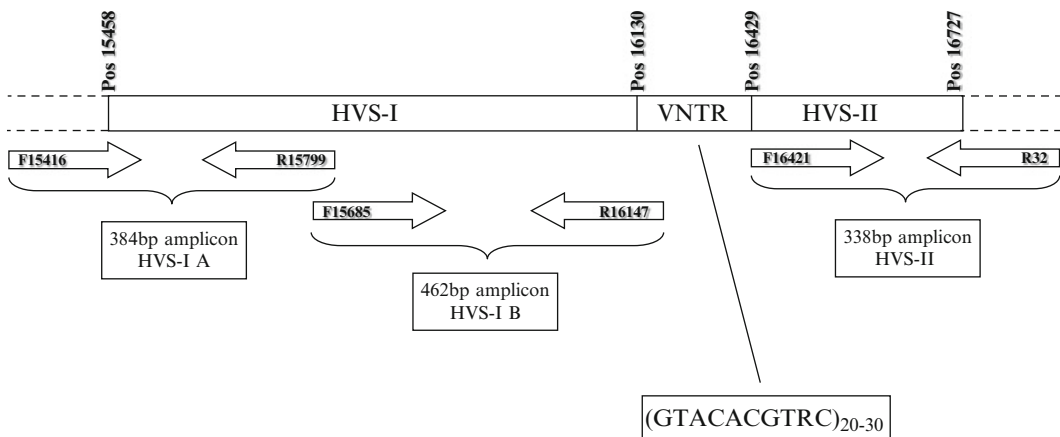
Fig. 1. Schematic representation of the hypervariable segments (HVS) I and II and the VNTR region of the canine mitochondrial DNA (mtDNA) control region. The amplification and sequencing strategy is illustrated by *arrows*, the resulting amplicons are shown in *boxes*. PCR and sequencing primers are noted inside the *arrows* (for primer nomenclature and sequences see Table 1). For better amplification success, HVS-I was divided into two shorter amplicons.

shafts, the amplification of the entire HVS-I region (672 bp) is not always successful. Therefore, two overlapping segments (HVS-I A and HVS-I B) with amplicon sizes of 384 and 462 bp are amplified and sequenced, respectively. DNA sequencing is based on the Sanger Big-dye deoxy-terminator method and the separation of the sequencing reaction products is performed on a conventional multicolor fluorescence detection capillary electrophoresis platform. Finally, software-aided base calling and sequence interpretation are addressed exemplarily.

2. Materials

The workflow described herein takes about 3 days time if performed in 8-h day shifts (see Note 1). Carry out all procedures at room temperature unless otherwise specified. Always use chemicals, including water, of very high purity grades (e.g., PCR grade water). Chemicals can be stored at room temperature unless otherwise specified. Typically, in-house prepared solutions can be stored up to half a year, but to minimize storage times, it is recommended to adjust the quantity produced to the current consumption. The use of factory-made buffers and solutions is preferable to an in-house production of the solutions in order to minimize risk of contamination and to improve quality consistency. For the same reasons, use sterile disposable consumables like solution containers, PCR tubes, pipette tips, etc.

2.1. Sample Preparation

1. Forceps.
2. Scissors.
3. Paper cleaning tissues.
4. 15% Hydrogen peroxide solution (H_2O_2): Store at 4–8°C. Prepare freshly at least every other day.

2.2. Lysis

1. 1 M Tris–HCl buffer pH 8.0.
2. 5 M NaCl solution.
3. 1 M $CaCl_2$ solution.
4. 10% SDS solution.
5. Hair extraction buffer: For 100 mL, add 1 mL 1 M Tris–HCl buffer pH 8.0, 2 mL 5 M NaCl solution, 100 μL 1 M $CaCl_2$ solution, 20 mL 10% SDS solution to 77 mL water.
6. Proteinase K (PK) solution: prepare a 20 mg/mL solution. To facilitate dissolving, incubate at 37° for 30 min, store at –20°C.
7. 1 M DTT solution, store at 4–8°C.
8. PCR grade water.

Table 1
Primer sequences used for amplification and sequencing of the two HVS of the canine mtDNA control region

Primer	Amplification segment	Primer sequence (5′–3′)
F15416	HVS-I A	CATCAGCACCCAAAGCTGAGA
R15799	HVS-I A	GTAAGAACCAGATGCCAGGTATA
F15685	HVS-I B	TCACCATGCCTCGAGAAACC
R16147	HVS-I B	ACGTGTACGTACGTGTACCYTAA AACTATA
F16421	HVS-II	TACACGTACGYACGCGC
R32	HVS-II	CGACTCATCTTGGCATTT

The primers are named according to the 5′-end with respect to the reference sequence (12). Strand designation is by forward (F) and reverse (R). The "Y" in the sequence of primers F16421 and R16147 indicate a mixture of "C" and "T"

2.3. Phenol/Chloroform Extraction

1. Ethanol absolute, store 50 mL aliquots intended for use during DNA extraction at –20°C.

2. Aqua Phenol pH 8 solution (toxic), store at 4–8°C for a maximum of 4 months.

3. Ready Red Chloroform-Isoamylalcohol mixture (toxic).

4. 3 M Na-Acetate-solution, pH 5.2.

5. 70% Ethanol.

6. PCR grade water.

2.4. Amplification

1. Advantage® cDNA Polymerase Mix and PCR Kit (Clontech Laboratories, Mountain View, CA, USA) including 1× BD Advantage 2 SA PCR Buffer and 1× BD Advantage 2 Polymerase Mix; store at –20°C (see Note 2).

2. Deoxynucleotide Triphosphate (dNTPs) solution mix: GeneAmp® dNTP Blend, 10 mM (Applied Biosystems, Foster City, CA, USA); store at –20°C.

3. Bovine Serum Albumin (BSA) solution (2.5 mg/mL), store at –20°C.

4. TLE-solution ($T_{10}E_{0.1}$): 10 mM Tris–HCl buffer pH 8.0 and 0.1 mM EDTA.

5. Primer: Each primer is diluted to 10 µM in TLE-solution, store at –20°C. For primer nomenclature and sequences, see Table 1.

6. PCR grade water.

2.5. Postamplification Sample Clean-Up	1. ExoSAP-IT for PCR Product Clean-up (USB, Cleveland, OH, USA). ExoSAP-IT is a mixture of the two hydrolytic enzymes, Exonuclease I and Shrimp Alkaline Phosphatase (SAP), for the removal of residual PCR primers and dNTPs.
2.6. Sequencing	1. BigDye® Terminator v1.1 Cycle Sequencing Kit and the BigDye® Terminator v1.1 and v3.1 5× Sequencing Buffer (both Applied Biosystems, Foster City, CA, USA).
	2. Primer (see Subheading 2.4).
	3. PCR grade water.
2.7. Postsequencing Sample Clean-Up	1. Illustra Sephadex G-50 Fine DNA Grade (GE Healthcare, UK).
	2. Multiscreen Column Loader 45 μL (Millipore, Billerica, MA, USA).
	3. Scraper for Multiscreen Column Loader (Millipore, Billerica, MA, USA).
	4. AcroPrep 96 Filter Plate (Pall Life Science, Ann Arbor, MI, USA).
	5. 96-Well plate, minimum 400 μL/well.
	6. MicroAmp 96-well Reaction Plate (Applied Biosystems, Foster City, CA, USA).
	7. PCR grade water.
2.8. Capillary Electrophoresis	1. 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).
	2. POP-6™ Polymer for 3100/3100-Avant™ Genetic Analyzers (Applied Biosystems, Foster City, CA, USA).
	3. ABI PRISM® 3100 Genetic Analyzer Capillary Array, 36 cm (Applied Biosystems, Foster City, CA, USA).
	4. MicroAmp 96-well Reaction Plate (Applied Biosystems, Foster City, CA, USA).
	5. MicroAmp Trays and Tray/Retainer Sets (Applied Biosystems, Foster City, CA, USA).
	6. 96-Well Plate Septa (Applied Biosystems, Foster City, CA, USA).

3. Methods

mtDNA analysis of challenging samples like hair shafts demands great care to avoid contamination. An overview of common practice to minimize contamination in forensic laboratories performing mtDNA is given in Note 3.

Mix all solutions before use.

Extraction blanks, negative and positive controls should be carried through the complete extraction, amplification, and sequencing processes for quality checks of the reagents and for contamination control.

3.1. Sample Preparation

1. Make sure that all reaction tubes are labeled unambiguously in order to avoid sample mix-up.

2. Put the functional parts of the forceps and scissors into the 15% H_2O_2 solution for at least 2 min and afterwards dry carefully with a clean paper tissue.

3. Cut the hair shaft into about 1 cm long pieces using the cleaned forceps and scissors and put 1–2 pieces of one single hair into a 1.5 mL sterile reaction tube that already contains 500 μL of hair extraction buffer (see Note 4). Continue directly with Subheading 3.2 or store the sample at 4°C.

3.2. Lysis

1. Add 500 μL hair extraction buffer, 20 μL PK solution, and 20 μL DTT solution.

2. Mix gently.

3. Incubate the tube at 56°C for at least 2 h using a thermal mixer adjusted to 650 rpm throughout the incubation period (see Note 5).

3.3. Phenol/Chloroform Extraction

1. Centrifuge the sample tubes for 30 s at about $16,300 \times g$ (see Note 6).

2. Add 500 μL of the Aqua Phenol pH 8 solution to the sample tube.

3. Mix gently.

4. Centrifuge for 10 min at $16,300 \times g$ to separate the aqueous phase and the phenol phase, the top phase being aqueous and containing the DNA.

5. Add 500 μL of the Ready Red Chloroform-Isoamylalcohol mixture to a new 1.5-mL reaction tube.

6. Transfer the upper (aqueous) phase from the sample tube into the reaction tube containing the Chloroform-Isoamylacohol using a pipette.

7. Mix gently.

8. Centrifuge for 10 min at $16,300 \times g$ forming two liquid phases, the top phase being aqueous and containing the DNA.

9. Add 50 μL of 3 M Na-Acetate pH 5.2 to a new 1.5-mL reaction tube.

10. Withdraw the upper (aqueous) phase to the new tube containing the Na-Acetate.

11. Mix gently.

12. Add 800 μL Ethanol absolute, cooled to –20°C.

13. Mix gently.

14. Cool at –20°C for at least 2 h.

15. Centrifuge for 30 min at $16,300 \times g$ to obtain an ethanolic supernatant and a DNA pellet (see Note 7).

16. Remove the Ethanol supernatant.

17. Add 1 mL of 70% Ethanol to the DNA pellet.

18. Mix gently.

19. Centrifuge for 20 min at $16,300 \times g$.

20. Remove the ethanolic supernatant.

21. Dry the samples for 30 min using a vacuum centrifuge.

22. Resuspend the DNA pellet in 50 μL of 10 mM Tris-buffer pH 8 overnight at 4°C or at 56°C for at least 2 h.

3.4. Amplification of the Canine mtDNA CR

The amplification strategy presented here (see Fig. 1) is primarily designed for the analysis of telogen hair and hair shaft samples as well as degraded blood and tissue samples (see Note 8). In the following, the entire amplification and sequencing processes will be described exemplarily for HVS-I A. For amplification, the primer combination F15416 and R15799 is used (see Note 9). Sequencing of the resulting amplicon is carried out by means of two separate sequencing reactions using F15416 and R15799 as sequencing primers, to obtain double strand sequencing results. In analogy to the procedure described for HVS-I A, the other segments of the CR are amplified and sequenced using the primers shown in Fig. 1.

1. Prepare a sample sheet that contains all important information you need for the PCR setup, e.g., plate position, sample name.

2. Use a UV light source for sterilization of the reaction tubes and plates used for amplification for at least 15 min.

3. Prepare an adequate volume of reagent PCR master mix (PCRMM) in a sterile reaction tube according to the intended number of PCR reactions (see Note 10). The PCRMM for a 25 μL PCR-assay contains per sample (final concentrations in brackets): 2.5 μL 1× BD Advantage 2 SA PCR Buffer, 2.5 μL BSA solution (0.25 mg/mL), 2 μL dNTPs (200 μM each dNTP), 0.75 μL of each primer F15416 and R15799 (300 nM), 0.5 μL 1× BD Advantage 2 Polymerase Mix. This results in 9 μL MM and a residual volume of 16 μL for the DNA sample or water per assay.

4. First pipette the residual amount of water into the wells of the PCR plate (where necessary), then add 9 μL PCRMM and

at last add the DNA extract. Pipette tips have to be changed on every new sample to prevent contamination during liquid transfers.

5. Cover the plate with a PCR sealing film.

6. Centrifuge for 30 s at about $1,840 \times g$ to remove air bubbles.

7. Carry out the PCR in a conventional thermal cycler using following protocol: initial denaturation at 95°C for 2 min followed by 39 cycles of 95°C for 15 s, 60°C for 30 s and 72°C for 45 s and a final step at 4°C forever.

3.5. Postamplification Sample Clean-Up

All amplicons (regardless of the primer combination used) are purified from residual primers and dNTPs using ExoSAP-IT.

1. Put 2 μL ExoSAP-IT into the well of a reaction plate for each sequencing reaction and add 5 μL of the PCR product.

2. Cover the plate with a PCR sealing film.

3. Centrifuge for 30 s at about $1,840 \times g$ to remove air bubbles.

4. Carry out the following temperature protocol in a conventional thermal cycler: incubation at 37°C for 15 min and enzyme inactivation at 80°C for 15 min.

3.6. Sequencing

Two sequencing reactions per sample are required using primers F15416 and R15799, respectively, to obtain double strand sequencing results for HVS-I A.

1. Prepare a sample sheet that contains all important information you need for the sequencing assay, e.g., plate position, sample name.

2. Prepare an adequate volume of reagent master mix (SeqMM) in a sterile reaction tube according to the intended number of sequencing reactions. The SeqMM for a 10 μL assay contains per sample (final concentrations in brackets): 2 μL BigDye® Terminator v1.1 Cycle Sequencing Kit, 2 μL BigDye® Terminator v1.1 and v3.1 5× Sequencing Buffer, 0.5 μL primer F15416 or R15799 (0.5 μM), and 0.5 μL water.

3. Put 5 μL SeqMM into each well of a reaction plate and add 5 μL of the purified amplification products (PCR product/ ExoSAP-IT mixture).

4. Cover the plate with a PCR sealing film.

5. Centrifuge for 30 s at about $1,840 \times g$ to remove air bubbles.

6. Carry out the following temperature protocol in a conventional thermal cycler: initial denaturation step at 96°C for 60 s followed by 30 cycles at 96°C for 15 s, 50°C for 5 s and 60°C for 4 min and a final step at 4°C forever.

3.7. Postsequencing Sample Clean-Up

All sequencing reaction products (regardless of the sequencing primer used) are purified from residual dye terminators using Sephadex G-50 Fine.

1. In order to load the Sephadex resin simultaneously into all wells of a 96-well plate, pour an excess of the resin onto the column loader.

2. Push the resin into all wells using the scraper and abrade the surface of the loader to the open end. The remaining Sephadex material can be stored and reused.

3. Place an empty AcroPrep 96 Filter Plate upside down on the top of the column loader. Push the plate against the end stop to align the columns to the plate wells.

4. Invert both the column loader and the AcroPrep 96 Filter Plate and tap on the backside of the loader in order to release the Sephadex powder to the wells.

5. Place the AcroPrep 96 Filter Plate on a 96-well plate and add 300 μL water to each well of the Filter Plate. Allow the powder to swell for at least 3 h at room temperature (see Note 11).

6. Centrifuge at $910 \times g$ for 5 min to rinse the water into the 96-well plate.

7. Discard the water in the 96-well plate.

8. Add 150 μL water to each well of the Filter Plate.

9. Centrifuge at 910 g for 5 min to rinse the water into the 96-well plate.

10. Discard the water in the 96-well plate.

11. Add 15 μL water to the sequencing reaction samples.

12. Place the Filter Plate on a 96-well plate. By choosing MicroAmp 96-well reaction plates which are compatible to the 3100 Genetic Analyzer, it is possible to continue directly with the capillary electrophoresis after the Postsequencing Sample Clean-up has been completed.

13. Pipette the sequencing products to the center of each well. Avoid touching the resin-columns with the pipette tip as this can decrease the separation performance of the column.

14. Centrifuge the plate/Filter Plate assembly at $910 \times g$ for 5 min to elute the sequencing products into the wells of the MicroAmp 96-well plate.

15. Discard the Filter Plate.

16. The sequencing products are now ready for capillary electrophoresis.

3.8. Capillary Electrophoresis on the Genetic Analyzer AB3100

1. Cover the plate containing the sequencing products with a 96-well plate septum and mount into a tray/retainer set.

2. Capillary electrophoresis is carried out on an ABI 3100 Genetic Analyzer using POP-6, 36-cm capillary arrays, and default instrument settings for sample loading and separation as recommended by the manufacturer.

3. Raw-data are analyzed with the Sequencing Analysis software Version 3.7 (Applied Biosystems, Foster City, CA, USA).

3.9. Sequence Data Analysis and Reporting

The data of the individual sequencing reactions are edited and aligned using the software Sequencher Version 4.7 (GeneCodes, Ann Arbor, MI, USA). If possible, confirm a successful analysis of a hair by a second independent analysis of the residual hair (shaft). The total range of analysis described herein covers the entire mtDNA CR excluding the VNTR repeat region. The mtDNA control region of the dog ranges from position 15458 to position 16727 including a 10 bp VNTR region located between position 16130 and 16429 (see Note 12). If the amplification primers are located inside the analyzed range of the sequence, the primer-binding sequence must be excluded, as it reflects the sequence of the primer and not the genuine sequence information. Each position of the analyzed mtDNA sequence should be covered at least twice by double strand analysis including the sequence information of the L-strand (light) and the H-strand (heavy) in order to identify sequencing artifacts, unambiguous base calling, and heteroplasmic positions. Haplotype nomenclature is adopted from the human mtDNA field (17, 18) and representing a difference-coded annotation with respect to the reference sequence U96639 (13). In Fig. 2, alignment and annotation for a canine mtDNA sequence to the reference sequence is shown and in Table 2 mtDNA haplotypes observed in an Austrian population sample are listed exemplarily.

Typical estimates to describe the discrimination power of mtDNA typing are the exclusion capacity (14, 19–21) and the match probability (15, 22). Population studies on dogs meeting high forensic standards are necessary for estimations of canine mtDNA haplotype frequencies and for calculating match probabilities. However, these basic requirements are rarely fulfilled (e.g., see ref. (13–16) and therefore the establishment of population data or even canine mtDNA databases is an essential task for future research to improve the statistical power of canine mtDNA analysis. Due to its maternal mode of inheritance, the exclusion capacity retrieved from mtDNA polymorphisms is generally lower compared to that achieved by nuclear DNA analysis. The exclusion capacity of the entire canine CR (with exception of the VNTR region) typically amounts between 0.93 and 0.95 (14, 23). In some instances, however, the ability to exclude an evidence sample and a reference sample is already sufficient to answer forensically relevant questions.
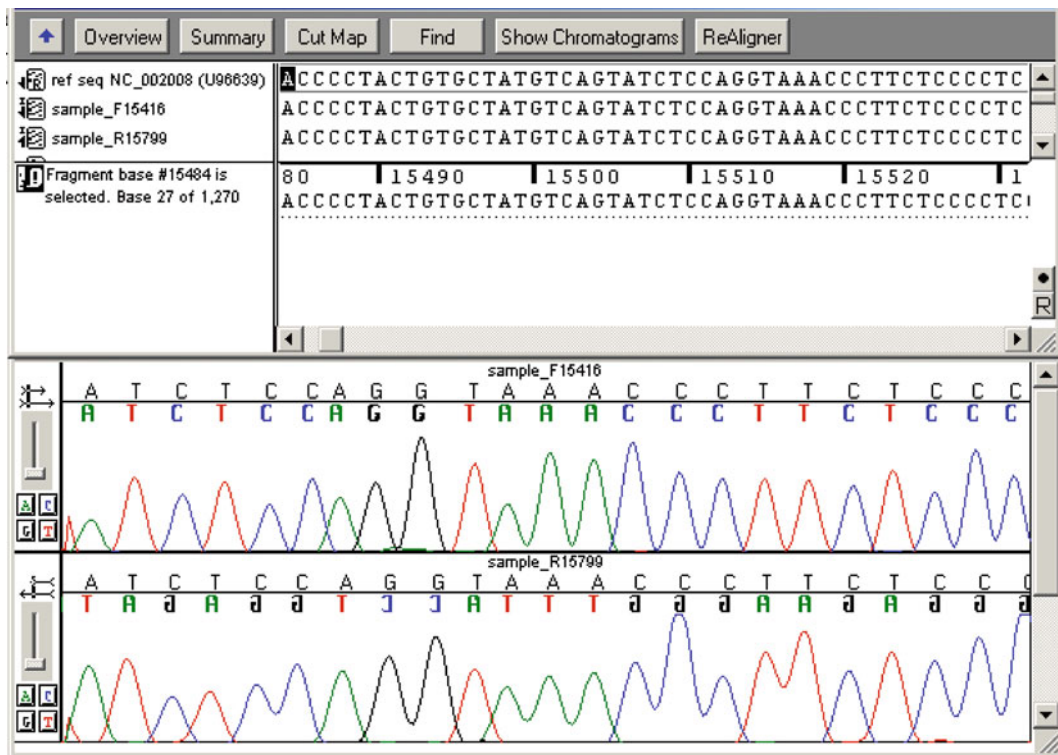
Fig. 2. Example of raw lane data aligned to the canine reference sequence NC_002008 (12).

4. Notes

1. The entire analysis from stain to profile will take at least 3 days:

   First day: Lysis (at least 2 h, often it is advisable to incubate the samples overnight) and DNA extraction (at least 4–5 h when performed by Phenol/Chloroform-Isoamylalcohol).

   Second day: Amplification process (approximately 2 h), post-PCR purification step (approximately 30 min), sequencing process (approximately 3 h), and postsequencing purification step (approximately 30 min).

   Third day: Electrophoretical separation (approximately 1 h per injection run) and analysis of the obtained sequences.

2. The Advantage cDNA Polymerase Mix includes KlenTaq-1 DNA polymerase (a 5′-exo-minus, N-terminal deletion of Taq DNA polymerase), a second DNA polymerase to provide 3′–5′ proofreading activity, and TaqStart® Antibody to provide automatic "hot start" PCR. In our experience, the Advantage cDNA Polymerase Mix provides reliable and efficient amplification results of canine mtDNA templates. If another DNA polymerase is intended to be used, the composition of the

Table 2

Observed polymorphic nucleotide positions of the canine mtDNA CR aligned to the reference sequence NC_002008, U96639 ([12]) in an Austrian dog population sample

The table reports, for each observed haplotype (HT) with its frequency (n), the variant nucleotides relative to the reference sequence (Ref Seq). A dot (.) indicates identity with the reference base. Insertion positions are marked with `.1`, `.2`, `.3`; `DEL` indicates a deletion; `R` is the IUPAC ambiguity code (A/G).

Reference sequence bases across the polymorphic positions (reading left to right; `|` marks grouped insertion/variant regions):

Ref Seq: `– C C G A C T T T T A C C A T T A T G T C T A T C C T C G C A | G | C C A T T A G C T T T T A | – – T T – C C | – C`

HT	n	Variant profile (positions differing from Ref Seq)
1	11	(identical to reference)
2	1	T . . . A
3	8	T
4	1	C . T
5	1	T T . T
6	20	A . C . T
7	1	A . G . C . T . . T
8	1	A . C . T . . . C
9	1	A . C . T . T
10	1	A . C . R . T
11	1	A . . C . T . . G . . T
12	1	A . C . T . . C . . T
13	2	G . A . C . C . T . . C
14	1	G . A . . T . . G . . T
15	1	G . A . . T . DEL . T . . T

Sequence alignment table. Dots (.) indicate identity with the reference sequence; letters indicate substitutions; DEL indicates a deletion.

Ref Seq	-CCGCACTTTACCATTATGTCTATCCTCGCA	G	CCATTAGCTTTA	---TT-CC	-C
16 1	. . . G . . . . A . A	.	. T	. . . .	. .
17 3	. . . G . T . . A	.	. . C	. . . .	. .
18 1	. . . G . T . . A	.	. . . .	. . . .	. .
19 1	. . . G . . C A . A C . . T . T	.	. T	. . . .	. C
20 1	. . C . T T . C A . . . . C . T C . A T DEL	.	T G . C . T C . C . . .	. . . .	. .
21 2	. . C C G . . . A . . . . T	.	T	. . C C	. T
22 13	. . C . G . . . A . . . . T	.	T	. . C C	. T
22 1	. . Y . G . . . A . . . . T	.	T	. . C C	. T
22 1	. . C . G . . . A DEL/A . . T	.	T	. . C C	. T
23 5	. . C . G . . . A . . . . T	.	. . . .	. . C C	. T
24 2	. . C . G . . . A . . . . T	.	. . . . C	. . C C	. T
24 1	. . C . G . . R . A . . . . T	.	. . . . C	. . C C	. T
25 1	. . C . G . . . A . . . . T	.	T	. . C C C . . .	DEL
26 4	. G . . . . A . . . . T	.	. . . .	. . . G	. T
27 3	. G . . . . A . . . . T	.	. . . . C	. . . .	. .
28 1	. G . . C . G . . . A . . . . T	.	. . . .	. . C C	. T
29 2	. T . T . C . . . G G . A . . . C . T C . T	.	T . G . . A T C	. . C C	. T
30 13	. T . T . C . . . G G . A . . . C . T C . T	.	T . G . . G A T C . . .	. . C C	. T
30 1	. T . T . C . . . G G . A . . . C . T C . T DEL/A	.	T . G . . G A T C . . .	. . C C	. T

Table 2
(continued)

The following table lists observed polymorphic nucleotide positions of the canine mtDNA control region. Positions are given as stacked (vertical) multi-digit numbers in the original; reference bases are shown in the "Ref Seq" row and haplotype substitutions are indicated relative to this reference.

HT	n	–	C	G	C	A	C	T	T	T	A	C	C	A	T	T	A	T	G	T	C	T	A	T	C	C	T	C	G	C	A	G	C	C	A	T	T	A	G	C	T	T	T	A	–	–	–	T	T	–	C	C	–	C
Ref Seq		–	C	G	C	A	C	T	T	T	A	C	C	A	T	T	A	T	G	T	C	T	A	T	C	C	T	C	G	C	A	G	C	C	A	T	T	A	G	C	T	T	T	A	–	–	–	T	T	–	C	C	–	C
31	1	.	.	.	T	.	T	.	C	.	.	T	.	.	.	.	.	.	G	G	.	A	.	.	C	.	T	C	.	.	T	.	.	T	.	G	.	.	G	A	T	C	.	.	.	.	.	T	.	.	.	.	.	.
32	2	.	.	.	T	.	T	.	C	.	.	.	.	.	.	.	.	.	G	G	.	A	.	.	C	.	T	C	.	.	T DEL	.	.	T	.	G	.	.	G	A	T	C	.	.	.	.	.	T	.	.	.	.	.	.
33	1	.	.	A	T	.	T	.	C	.	.	.	.	.	.	.	.	.	G	G	.	A	.	.	C	.	T	C	.	.	T	.	.	T	.	G	.	.	G	A	T	C	.	.	.	.	.	T	.	.	.	.	.	.
34	1	.	.	A	T	.	T	.	C	.	.	.	.	.	.	.	.	.	G	G	.	A	.	.	C	.	T	C	.	.	T DEL	.	.	T	.	G	.	.	G	A	T	C	.	.	.	.	.	T	.	.	.	.	.	.
35	2	.	.	.	T	.	T	.	C	.	.	.	.	.	.	.	.	.	G	.	.	C	.	T	.	.	C	.	T	.	T	DEL	T	G	.	.	.	C	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.
36	3	.	.	.	T	.	T	.	C	.	.	.	.	.	.	.	.	.	G	.	.	C	.	.	.	.	C	.	T	.	T	.	.	T	.	G	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	T
37	4	.	T	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
38	7	.	T	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
38	1	.	T	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
39	1	.	T	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
40	1	C	.	.	T	.	C	.	.	.	.	.	.	.	.	.	.	.	G	G	.	A	.	.	C	.	T	C	.	.	T	.	.	T	.	G	.	.	G	A	T	C	.	.	.	.	.	T	.	.	.	.	.	.

Dots indicate no differences to the reference sequence, a base call shows the substitution at a distinct position, DEL indicates a deletion, insertions are designated by numbers after a decimal point (".1," ".2," ".3"). Point heteroplasmy was designated with an International Union of Biochemistry code (mixture of purine bases R, a mixture of pyrimidine bases Y)

Observed polymorphic nucleotide positions of the canine mtDNA CR aligned to the reference sequence NC_002008, U96639 [12] in an Austrian dog population sample

HT Haplotype designation; *n* number of samples

reagents and the thermocycler protocol has to be adopted according to the characteristics of the selected enzyme (e.g., by using Taq polymerase, the initial denaturation step has to be increased from 2 to 11 min).

3. It is imperative to separate pre-PCR and post-PCR areas physically, usually by separate laboratories. Generally perform frequent cleaning procedures with bleach or UV irradiation in the pre-PCR working space and always clean the lab bench surfaces and hoods with the 15% H_2O_2 solution, bleach, or another appropriate disinfectant before preparing forensic samples for mtDNA analysis.

    Disposable sterile (DNase and RNase-free) PCR reagent container and tubes as well as aerosol-resistant filter pipette tips should be used.

    Reagents should be carefully prepared to minimize contamination. Usually a safety cabinet that is equipped with UV light is used for setting up PCR. PCR consumables (except primers, dNTPs, and Taq DNA Polymerase) are UV irradiated for at least 15 min prior to use.

    Use protective clothing and always wear gloves during sample handling or when working in the PCR area. Change gloves frequently.

4. If a hair root is visible, carry out an independent analysis (possibly by STR) or store separately.

5. Make sure that the entire hair shaft is completely immersed into the hair extraction buffer. If necessary, centrifuge the sample tube to move the hair to the bottom.

6. After lysis, the hair should be completely dissolved, therefore after centrifugation no residual hair fragments should be visible at the bottom of the tube. If parts are still visible, add 15 μL of PK and DTT, and incubate at 56°C for another 30 min using a thermal mixer adjusted to 650 rpm.

7. Because of the very small quantities of DNA from hair, a pellet is often not visible. The exact position of the pellet must therefore be deduced from the orientation of the tube during centrifugation. The label of the reaction tube can function as a good indicator if the position of the tube in the centrifuge is marked. The exact location of the DNA pellet is important for prevention of its unintentional removal together with the ethanol supernatant. For the removal of ethanol, it is advisable to move the pipette tip to a position opposite for the position of the DNA pellet at the bottom of the tube.

8. Depending on mtDNA amount and quality, different amplification strategies can be applied. For reference samples that contain high amounts of high-quality DNA such as blood or

Table 3
List of animal species that gave PCR products with the PCR primers for HVS-I A

Animal species	Ref Seq	Sequence of the F-primer binding region	Location	Pos	Amplicon length
Dog	U96639	CATCAGCACCCAAAGCTGAGA GTAAGAACCAGATGCCAGGTATA	D-loop	15416–15799	384
Wolf	NC_009686	CATCAGCACCCAAAGCTGAAA GTAAGAACCAGATGCCAGGTATA	D-loop	15417–15800	384
Fox	AM181037	CGTCAGCACCCAAAGCTGAAA GTAAGAACCAGATGCCAGGTATA	D-loop	15418–15778	361
Racoon dog	D83614	------ GTAAGAACCAGATGCCAGGTATA	D-loop	–332	At least 332
Pine marten	AF336952	CATCAGCACCCAAAGCTGACA -----ACCAGATGCCAGGTATA	D-loop	112–445	334
Ferret/polecat	AY962044	CATCAGCACCCAAAGCTGATA GTAAGAACCAGATGCCAGATATA	D-Loop	153–496	344
Cat	U20753	TATCAGCACCCAAAGCTGAAA GTAAGAACCAGATGCCAGGTATA	D-loop	16272–16949	678
Lynx	AY034816	CATCA : CTCCCAAAGCTGAAA GTAGGAACCAGATGCCAGGTATA	D-loop	1–698	698
Horse	NC_001640	CATCAACACCCAAAGCTGAAA GAAAGAACCAGATGCCAGGTATA	D-loop	15427–15856	430
Donkey	NC_001788	CATCAACACCCAAAGCTGAAA GAAAGAACCAGATGCCAGGTATA	D-loop	15422–15851	430
Hare	NC_004028	CATCAGCACCCAAAGCTGAAA GTAAGAACCAGATGCCAGTTATA	D-Loop	15398–15930	533
Orangutan	AC150686	CTAAGAACCCAAAGCTGAGA CTAAGAACCAGATGCCAGGTATA	Chromosome 7[a]	103009–103544	536
Gorilla	AC150686	CATCAACACCCAAAGCTGAGA CTAAGAACCAGATGCCAGGTATA	Chromosome 7[a]	103009–103544	536
Chimpanzee	AC150686	CATCAACACCCAAAGCTGAGA CTAAGAACCAGATGCCAGGTATA	Chromosome 7	103009–103544	536
Human	AC099654	CATCAACACCCAAAGCTGAGA CTAAGAACCAGATGCCAGGTATA	Chromosome 7	47212–47747	536

Sequence differences of the primer-binding region compared to the primer sequences (dog) of F15416 and R15799 are indicated in *gray*. The accession numbers of the reference sequences of each animal species are listed. The approximate product length and basepair positions are listed as well as the chromosomal location

----- No sequence information found; : Deletion

[a] BLAST search resulted in chimpanzee chromosome 7

saliva, the entire CR can be amplified in one PCR reaction, e.g., by using the primer combination F15416 and R32 (see Fig. 1). In addition, the cycle number can be reduced from 39 to 35.

9. The primers used for the amplification of HVS-I A (F15416 and R15799) also amplify parts of the mtDNA control region of other animal species, e.g., cat and horse (see Table 3). DNA from humans and other primates gave 536 bp amplicons, both located on chromosome 7 (BAC clone RP11-1217F2 Access. Nr. AC099654.5 and Pan troglodytes BAC clone CH251-549B3 AC150686.2) (see Table 3). A similar primer combination was used by Nakamura et al. (24) for forensic species identification based on size variation of mtDNA regions. The primer combinations for the other two CR segments give no amplification products with DNA of other tested animal species.

10. If possible, prepare a fresh PCRMM to avoid too long storage times in the freezer, as the quality and the performance of the PCR will otherwise be decreased. Allow an additional volume of 5–10% to compensate for pipetting errors.

11. As Sephadex is supplied as a dry powder, it must be allowed to swell in water before use. Swelled Sephadex plates can be stored overnight in a fridge. Add water only to these wells that are intended for use, remaining wells can be used later.

12. Variation within the VNTR region is rarely used in canine forensics (25, 26). In contrast to our reporting strategy, the VNTR region is included by Pereira et al. (13).

Acknowledgment

References

1. Eichmann C, Berger B, Parson W (2004) A proposed nomenclature for 15 canine-specific polymorphic STR loci for forensic purposes. Int J Legal Med 118: 249–266

2. Eichmann C, Berger B, Reinhold M et al (2004) Canine-specific STR typing of saliva traces on dog bite wounds. Int J Legal Med 118(6): 337–342

3. Hellmann AP, Rohleder U, Eichmann C, et al (2006) A proposal for standardization in forensic canine DNA typing: allele nomenclature of six canine-specific STR loci; J Forensic Sci 51: 274–281

4. Berger C, Berger B, Parson W (2009) Canine DNA Profiling in Forensic Casework: The Tail Wagging the Dog. Forensic Sci Rev 21(1): 1–13

5. Van Asch B, Alves C, Gusmão L et al (2009) A new autosomal STR nineplex for canine identification and parentage testing. Electrophoresis 30(2): 417–23

6. Tom BK, Koskinen MT, Dayton M et al (2010) Development of a nomenclature system for a canine STR multiplex reagent kit. J Forensic Sci 55(3): 597–604

7. Zenke P, Egyed B, Zöldág L et al (2010) Population genetic study in Hungarian canine populations using forensically informative STR loci. Forensic Sci Int: Genetics doi 10.1016/j.fsigen.2010.03.013

8. Parson W, Parsons TJ, Scheithauer R et al (1998) Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. Int J Legal Med 111: 124–132

9. Melton T, Dimick G, Higgins B et al (2005) Forensic mitochondrial DNA analysis of 691 casework hairs. J Forensic Sci 50(1): 73–80

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. Berger C, Parson W (2009) Mini-midi-mito: Adapting the amplification and sequencing strategy of mtDNA to the degradation state of crime scene samples. Forensic Sci Int: Genetics: 149–153

12. Kim KS, Lee SE, Jeong HW et al (1998) The complete nucleotide sequence of the domestic dog (Canis familiaris) mitochondrial genome. Mol Phylogenet Evol 10(2): 210–220

13. Pereira L, Van Asch B, Amorim A (2004) Standardisation of nomenclature for dog mtDNA D-loop: a prerequisite for launching a Canis familiaris database. Forensic Sci Int 141(2-3): 99–108

14. Eichmann C and Parson W (2007) Molecular characterization of the canine mitochondrial DNA control region for forensic applications. Int J Legal Med 121(5): 411–16

15. Gundry RL Allard MW Moretti TR et al (2007) Mitochondrial DNA analysis of the domestic dog: control region variation within and among breeds. J Forensic Sci 52(3): 562–72

16. Desmyter S, Comblez S (2009) Belgian dog mitochondrial DNA database for forensics. Forensic Sci Int: Genetics Supplement Series 2(1): 286–287

17. Anderson S, Bankier AT, Barrell BG et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290: 457–465

18. Andrews RM, Kubacka I, Chinnery PF et al (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23: 147

19. Aaspollu A, Kelve M (2003) The first criminal case in Estonia with dog's DNA data admitted as evidence. Proceedings - Int Congress Series 1239: 847–851

20. Savolainen P, Rosen B, Holmberg A, et al (1997) Sequence analysis of domestic dog mitochondrial DNA for forensic use. J Forensic Sci 42(4): 593–600

21. Savolainen P, Lundeberg J (1999) Forensic evidence based on mtDNA from dog and wolf hairs. J Forensic Sci 44(1): 77–81

22. Halverson JL, Basten C (2005) Forensic DNA identification of animal-derived trace evidence: tools for linking victims and suspects. Croat Med J 46: 598–605

23. Angleby H, Savolainen P (2005) Forensic informativity of domestic dog mtDNA control region sequences. Forensic Sci In 154(2–3): 99–110

24. Nakamura H, Muro T, Imamura S et al (2009) Forensic species identification based on size variation of mitochondrial DNA hypervariable regions. Int J Legal Med 123: 177–184

25. Fridez F, Rochat S, Coquoz R (1999) Individual identification of cats and dogs using mitochondrial DNA tandem repeats? Sci Justice 39: 167–171

26. Savolainen P, Arvestad L, Lundeberg J (2000) A novel method for forensic DNA investigations: repeat-type sequence analysis of tandemly repeated mtDNA in domestic dogs. J Forensic Sci 45: 990–999

Part III

Advances in Microchip Capillary Electrophoresis

Chapter 24

Integrated Sample Cleanup and Microchip Capillary Array Electrophoresis for High-Performance Forensic STR Profiling

Peng Liu, Susan A. Greenspoon, Stephanie HI Yeung, James R. Scherer, and Richard A. Mathies

Abstract

Microfluidics has the potential to significantly improve the speed, throughput, and cost performance of electrophoretic short tandem repeat (STR) analysis by translating the process into a miniaturized and integrated format. Current STR analysis bypasses the post-PCR sample cleanup step in order to save time and cost, resulting in poor injection efficiency, bias against larger loci, and delicate injection timing controls. Here we describe the operation of an integrated high-throughput sample cleanup and capillary array electrophoresis microsystem that employs a streptavidin capture gel chemistry coupled to a simple direct-injection geometry for simultaneously analyzing 12 STR samples in less than 30 min with >10-fold improved sensitivity.

Key words: Capillary electrophoresis, Forensic human identification, Lab-on-a-chip, PCR cleanup, Short tandem repeat analysis

1. Introduction

Driven by the awareness of the power of DNA for human identification and the increasing sophistication in DNA profiling and evidence collection, the number of samples submitted for short tandem repeat (STR) analysis continues to rise dramatically, resulting in a nationwide backlog. In addition, forensic scientists often face the challenges of analyzing evidence with a wide range in DNA quantity and quality. A faster, cheaper and better analysis is needed to address these challenges (1–3). Microfabrication technology has the potential to significantly advance STR analysis by translating the entire process into a nanoliter scale in a miniaturized and integrated microfluidic format (4–6). Such an integrated system will
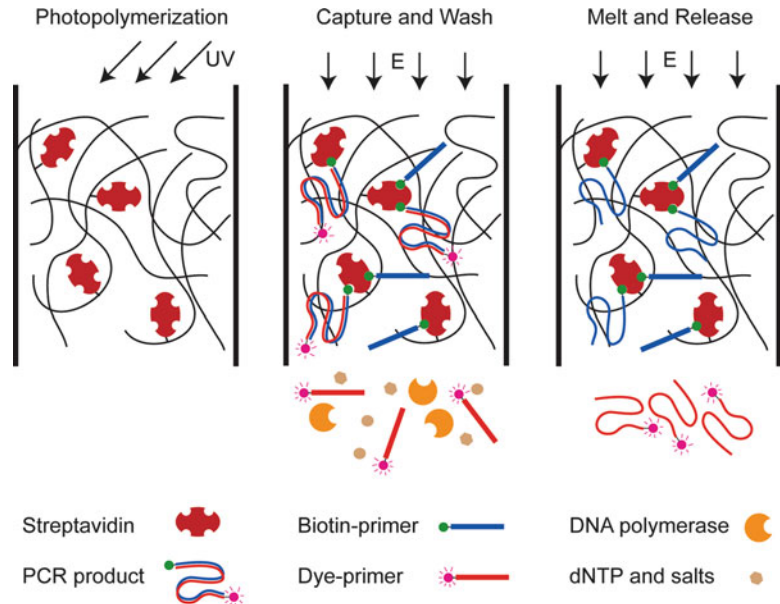
Fig. 1. Schematic of the streptavidin capture method for post-PCR cleanup and concentration of forensic PCR products. STR samples are generated with one primer labeled with fluorescent dye and one with biotin. The double-strand DNA products are electrophoresed through the photopolymerized cross-linked polyacrylamide gel network where DNA binds to the streptavidin and unbound reactants are washed off. The fluorescently labeled DNA strands are thermally released for electrophoresis.

decrease the reagent and time consumption, make the process more automated and robust, and reduce manual sample handling to minimize the risk of sample mix-up and contamination.

Current STR analysis employs a direct injection from high-salt PCR samples without cleanup for capillary electrophoresis (CE), resulting in poor injection efficiency (<1%) and bias against larger loci. Additionally, the cross injection geometry in conventional CE microchips requires delicate timing optimization and electric field balance. To address these issues, numerous post-PCR purification, concentration, and separation methods, such as membrane filtration (7, 8), sample stacking (9, 10), and sample extraction (11), have been explored. Although these approaches have demonstrated sensitivity improvement, the typically delicate operations are incompatible with high-throughput on-chip integration.

Researchers at the UC Berkeley have developed an integrated STR sample cleanup and separation microdevice and method that employs a streptavidin capture gel chemistry coupled to a simple direct-injection geometry (12). As illustrated in Fig. 1, a streptavidin-polyacrylamide capture gel is immobilized in a microchannel
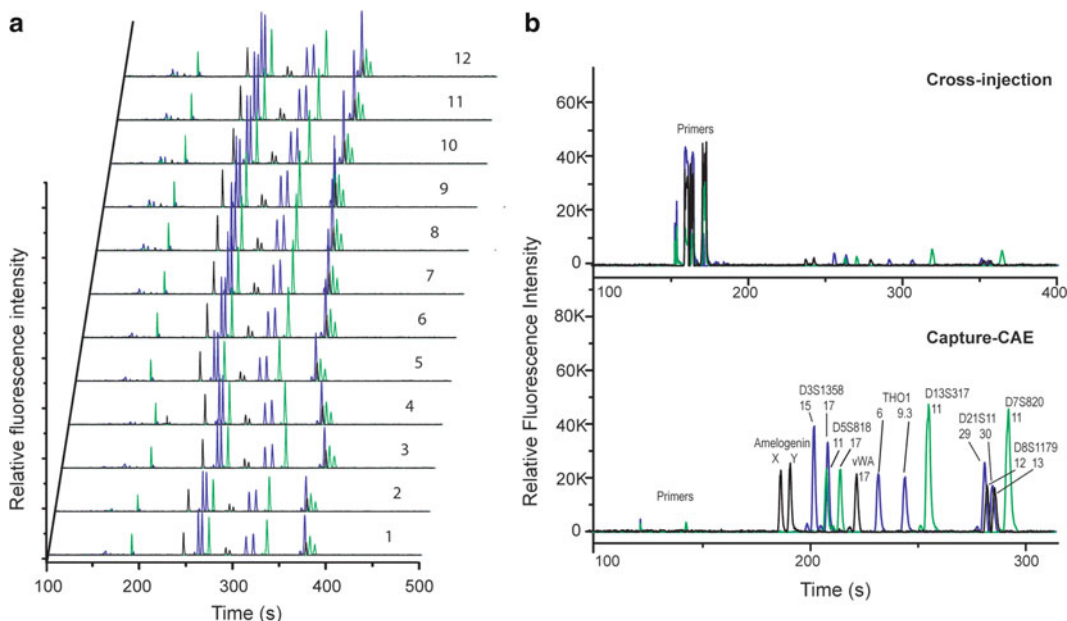
Fig. 2. (**a**) STR profiles of 0.15-ng standard DNA obtained from the 12-lane capture-CAE microdevice. All the traces from the 12 lanes are plotted on the same signal intensity scale. The average percentage standard deviation of the allele signal-to-noise (S/N) ratios is 18.8% and the average standard deviation of the allele migration time is 9.8 s. (**b**) Comparison of STR profiles obtained using the capture inline injection and the conventional cross injection methods under the same conditions. The signal intensities of STR samples amplified from 0.15 ng of 9948 standard DNA are improved > 10-fold. (Reproduced with permission from Ref. (13)).

by photopolymerization. Biotin-labeled double-stranded DNA products amplified by PCR using one biotin and one fluorescence-labeled primer are electrophoresed through the capture gel where DNAs are bound efficiently via the biotin-streptavidin interaction, while unbound materials are washed off. The fluorescently labeled DNA strands are thermally released for a high quality, quantitative electrophoretic injection. Compared to conventional microchip CE with a cross-injector, the fluorescence intensity was improved 14–19 fold for 9-plex STR products. To improve the throughput of this method, a 12-lane capture-CAE microsystem that utilizes the inline capture injection process with an improved automated process in a radial microfabricated capillary array electrophoresis (μCAE) format was successfully developed (13). As shown in Fig. 2, the process of analyzing 12 STR samples required no manual intervention and can be completed in only 30 min. A comparison between capture inline injection and conventional cross injection demonstrated at least 10-fold improvement in sensitivity. This microsystem provides a robust and sensitive platform for routine as well as low copy number (LCN) and degraded DNA analysis.

In this chapter, we are presenting the operation protocol of this integrated 12-lane capture-capillary array electrophoresis microsystem developed in the UC Berkeley-Virginia Department of Forensic Science collaboration, including microchannel and STR sample preparation, photopolymerization of capture gels, microchip setup, on-chip sample cleanup and capillary electrophoresis, and data analysis.

2. Materials

2.1. Capture-CAE Microsystem

1. A 12-lane capture-CAE microdevice: the microdevice is fabricated on 100-mm diameter glass wafers using published procedures (see Note 1). As shown in Fig. 3, a total of 12 separation channels are arranged radially around a common anode well on a 100-mm-diameter glass microchip. Adjacent pairs of channels are grouped into doublet structures that share a cathode and a waste well. Each separation channel has a double-T junction with an offset of 500 μm between the sample and waste arms for PCR product cleanup, concentration, and inline injection. The separation channels are 120 μm wide, 40 μm deep, and have an effective separation length of 10 cm.
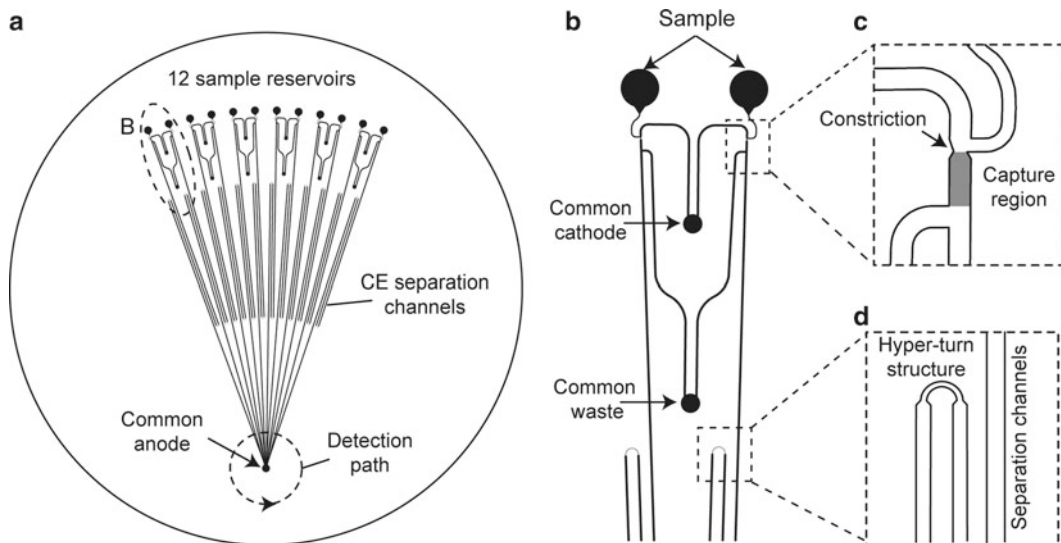
Fig. 3. Schematic of the capture-CAE microdevice. (**a**) A total of 12 separation channels with a common anode are arranged on a 4″ glass wafer. (**b**) Each doublet contains two capture gel inline injectors and two sample wells sharing one cathode and one waste well. (**c**) Expanded view of the gel capture region. (**d**) Expanded view of the hyper-turn structure in the separation channels. (Reproduced with permission from Ref. (13)).
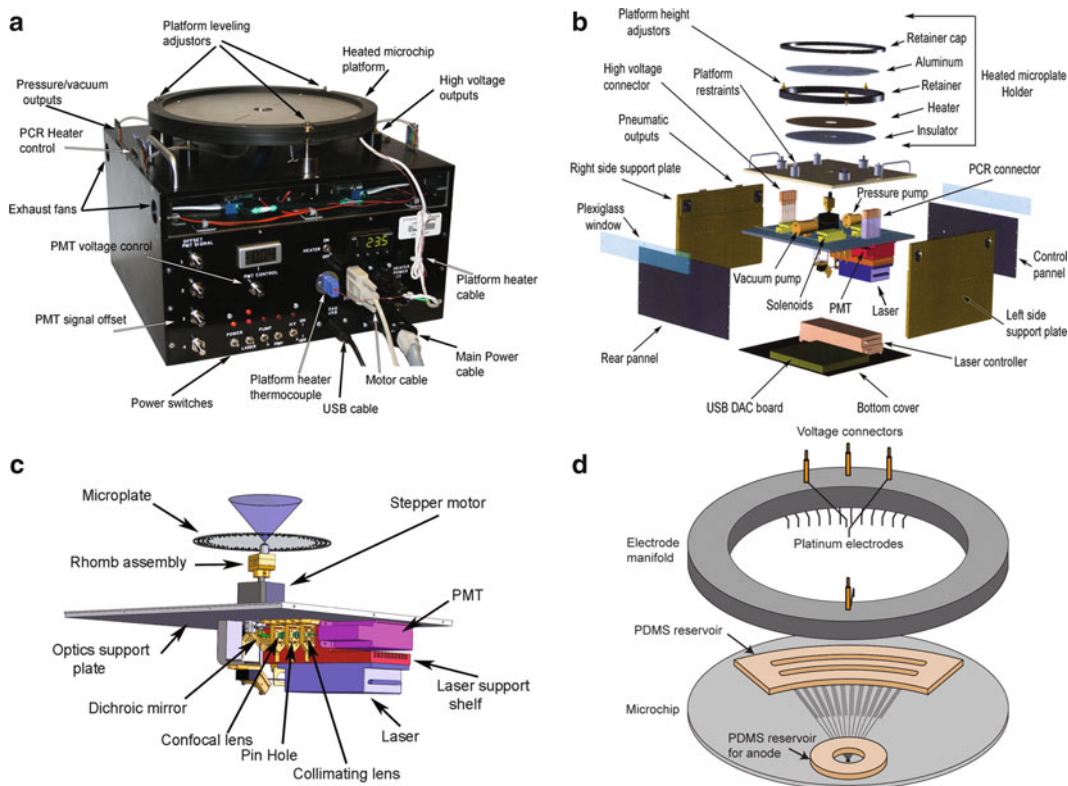
Fig. 4. Illustration of the microchip control and detection system. (**a**) Photograph of the Multi-channel Capillary Array Electrophoresis Portable Scanner (McCAEPs). (**b**) Exploded view of the instrument. The major structure of the instrument includes the detection platform, a main support plate on which the optical and pneumatic components are installed, two side support plates, and a bottom plate. (**c**) Schematic of the four-color confocal detection system with a rotary objective. (**d**) Schematic of the electrode manifold, PDMS reservoirs, and the microchip. (Reproduced with permission from Ref. (15)).

2. Four-color detection scanner: the Multi-channel Capillary Array Electrophoresis Portable Scanner (McCAEPs), shown in Fig. 4a–c, is employed for the control and detection of the microchip (see Note 2).

3. Two PDMS (polydimethysiloxane) reservoirs: as shown in Fig. 4d, one contains two cut-through arc slots for defining continuous buffer reservoirs of the cathode and the waste, the other has a 1-cm-diameter hole for the anode (see Note 3).

4. An electrode manifold: as shown in Fig. 4d, the manifold contains Pt electrodes that are positioned within the reservoirs on the microchip for the application of high voltages during electrophoresis.

5. High-pressure filling station (see Note 4).

2.2. Microchip Coating Components

1. Coating solution: polyDuramide (MW ~ 2,000,000) is the trade name for poly-*N*-hydroxyethylacrylamide. It is prepared at a concentration of 0.25% (w/v) in DI water: for example, dissolve 50 mg of polyDuramide in 20 mL of DI water (see Note 5).

2. 1 M HCl solution: the concentrated 37.5% HCl is 12.2 M, dilute to 1 M HCl by DI water.

3. 1 M NaOH.

4. Acetonitrile.

5. Gel Slick® glass plate coating solution (Lonza, Rockland, ME, USA).

2.3. Streptavidin Capture Gel Components

1. 30% Bis-acrylamide solution (19:1) (see Note 6).

2. 8 M urea in DI water.

3. Tris TAPS EDTA (TTE) buffer: the 10× TTE buffer contains 500 mM Tris-base, 500 mM TAPS acid, and 10 mM Na_2EDTA. Store the 10× TTE stock solution at 4°C. When use, dilute it to 1× by DI water.

4. Streptavidin acrylamide (Invitrogen, Carlsbad, CA, USA): 1 mg of powder.

5. Riboflavin: prepare a 0.01% (w/v) stock solution in DI water and store it at room temperature in dark (see Note 7).

6. *N,N,N′,N′*-Tetramethylethylenediamine (TEMED): store at 4°C. When ready for use, dilute TEMED to 50% (v/v) with DI water (see Note 8).

2.4. UV Exposure System

1. 5% linear polyacrylamide gel (LPA) in 1× TTE (see Note 9).

2. Photomask: the glass photomask produced by Photo Sciences Inc. (Torrance, CA, USA) features clear exposure windows and two alignment markers in the positions corresponding to the capture inline injectors and the markers on the microchip. Three vacuum access holes are diamond-drilled on the mask.

3. UV exposure system: a UV exposure setup installed on a Nikon inverted microscope (TE2000U) with a 200-W mercury arc lamp is employed for photopolymerization. The aluminum vacuum stage connects to an external vacuum line and is mounted on the microscope stage. The photomask featuring clear exposure windows is placed underneath the microchip and visually aligned to the microchannel under the microscope. The assembly is secured on the stage via vacuum access holes on the photomask. The UV light travels from the microscope objective up through the exposure windows and initiates polymerization in only the unmasked microchannel regions (see Note 10).

2.5. Sample Preparation

1. STR typing kits: a 9-plex autosomal STR typing system including amelogenin for sex typing and eight Combined DNA Index System (CODIS) core STR loci (D3S1358, TH01, D21S11, D5S818, D13S317, D7S820, vWA, and D8S1179) is employed for this system. To enable the capture cleanup, the primer pair of each locus has one primer labeled with fluorescent dye and the other labeled with biotin (see Note 11).

2. A biotin-modified internal lane standard (see Note 12).

3. Hi-Di™ formamide (Applied Biosystems, Foster City, CA, USA): aliquot formamide to small amounts (200 μL in each eppendorf tube) and store at –20°C. Use a fresh tube each time for sample preparation (see Note 13).

2.6. Microchip Preparation and Capillary Electrophoresis

1. LPA for electrophoresis: 5% LPA with 8 M urea in 1× TTE (see Note 9).

2. 2-propanol.

2.7. Data Analysis

1. BaseFinder 6.0 (see Note 14).

2. A data conversion program for the MeagBACE™ Fragment Profiler software (see Note 15).

3. MegaBACE™ Fragment Profiler 1.2 (GE Healthcare, Piscataway, NJ, USA).

3. Methods

Carry out all procedures at room temperature unless otherwise specified.

3.1. Microchip Coating

1. Fill the channels with 1 M NaOH, let the chip sit at 65°C for 15 min, and then empty the channels followed by rinsing with DI water.

2. Fill the channels with 1 M HCl and flush ~100 μL, let the chip sit at room temperature for 15 min, and then remove HCl followed by channel rinsing with DI water.

3. Load the polyDuramide coating solution into the channels, and let the chip sit for 15 min (see Note 16).

4. Remove coating solution and rinse the channels with DI water followed by acetonitrile to dry the channels completely.

5. Apply Gel Slick® solution onto a dry paper towel, and spread Gel Slick® evenly over the entire surface of the chip using a circular and overlapping motion. Allow the coating to dry completely, and wipe the chip surface again with soft paper towels to remove stains (see Note 17).

3.2. Preparation of Photopolymerization Solution

1. To prepare a 500 μL of photopolymerization solution, sequentially add 35.5 μL DI water, 250 μL 8 M Urea, 50 μL 10× TTE, and 83.3 μL 30% bis-acrylamide into an opaque 2-mL scintillation vial with Teflon closure.

2. Close the vial tightly and insert two syringe needles through the Teflon closure. One needle is inserted into the solution for bubbling nitrogen; the other is above the solution as a vent. Sparge the solution with a N_2 flow rate of 5 bubbles/s for 10 min (see Note 18).

3. Add 50 μL of DI water into the tube of streptavidin acrylamide. Slowly pipette the water 2–3 times to dissolve the streptavidin acrylamide to a concentration of 20 μg/μL (see Note 19).

4. Stop sparging the solution. In the dark room, mix the 50 μL streptavidin acrylamide solution, 30 μL 0.01% riboflavin, and 1.25 μL 50% TEMED together in a tube, and then inject the mixture into the vial through the Teflon closure using a syringe. Gently mix the solution, and store in dark at 4°C (see Note 20).

3.3. UV Exposure for Photopolymerization

1. Stack the photomask and the capture-CAE microchip on the aluminum vacuum stage installed on the inverted microscope. Turn on the vacuum to hold the mask and the chip together, and then align the microchip with the photomask under the microscope with the aid of alignment markers (see Note 21).

2. Take 100 μL of the photopolymerization solution out of the vial using a syringe with a needle. Fill the channels with the solution via capillary action. Evacuate excess solution from each well and replace with a drop of 5% LPA gel (see Note 22).

3. Move the exposure window of the photomask to the center of the microscope view. Open the UV light shutter to initiate the gel photopolymerization. Every 5 min, move the microscope stage to bring the next exposure window to the center of the view for exposure.

4. Once the exposure of all 12 lanes is finished, close the UV light shutter and turn off the mercury lamp. Evacuate the gel in the wells, and replace the photopolymerization buffer in the channels with 1× TTE buffer using vacuum.

3.4. DNA Sample Preparation

1. Amplify DNA samples using biotin-labeled STR typing kits on a PCR thermal cycler.

2. Add 1 μL of each PCR product into a loading cocktail containing 0.5 μL Hi-Di formamide, 2.5 μL DI H_2O, and 1 μL internal lane standard (see Note 23).

3.5. Microchip Preparation

1. Carefully vacuum out any solution left on the chip surface without emptying the channels, and clean the chip surface with DI water.

2. Soak the PDMS buffer reservoirs in 2-propanol, and then blow-dry the PDMS using a Nitrogen gun. Align and attach the PDMS reservoirs onto the microchip (see Note 24).

3. Fill the channel with 5% LPA with 8 M urea from the common anode using the high-pressure filling station until the gel coming out from the waste wells.

4. Manually fill the channels above the capture gel plugs with 5% LPA with 8 M urea from each common cathode well until the gel reaches the samples wells using a syringe with a pipette tip. The opening of the tip is cut to fit into the cathode wells for applying pressure.

5. Evacuate the excessive gel out from the sample wells using vacuum, and then load 1.5 µL prepared DNA samples into each sample well.

3.6. Capture-CAE Microchip Operation

1. Transfer the microchip onto the stage of the detection instrument and turn on the vacuum to hold the chip in place.

2. Fill the cathode, the waste, and the anode reservoir with 1× TTE.

3. Place the electrode manifold on the top of the microchip, and connect the electrodes to high-voltage power supplies.

4. Turn on the detection instrument and start the control program of the scanner. Level the stage to bring all the separation channels into focus and define the positions of the channels (see Note 25).

5. Inject the samples from the sample wells toward the waste wells for 10 min by applying an electric field of 25 V/cm from the sample to the waste wells while floating the cathode and anode wells. PCR products are bound efficiently to the capture gel via the biotin-streptavidin interaction to form a concentrated plug (see Note 26).

6. After capture, apply an electric field of 25 V/cm from the cathode to the waste to wash all the uncaptured sample contents through the gel to the waste for 5 min.

7. To clean the channels above the capture gel, apply the same electric field from the cathode to the sample. To clean the separation channels, use the same electric field to wash any unbound materials from the anode toward the waste.

8. Stop all the voltages to the microchip, and set the temperature of the detection stage of the instrument to 67°C. Once the stage temperature reaches the setting point, wait one more minute to equilibrate the setup.

9. Switch the voltages on each well to: 2,500 V to the anode, 200 V to the sample and the waste, and ground to the cathode wells. The resulting electric field is about 150 V/cm in the separation channels for electrophoresis. Start the detection system to record data (see Note 27).

10. After electrophoresis is completed in about 20 min, stop the high voltages and save the data. Take off the microchip from the stage and shut down the instrument.

11. Vacuum out the samples and buffers in the reservoirs and push out the sieving matrix and the capture gel plugs in the channels from the common anode well using the high-pressure filling station (see Note 28).

12. Thoroughly rinse the channels with DI water using the filling station. Dry the channels completely using acetonitrile. Store the microchip in a tightly sealed wafer tray (see Note 29).

13. Soak the electrode manifold and the PDMS buffer reservoirs into 2-propanol and ultrasonicate for 15 min to eliminate any sample residues. Rinse them with DI water and blow-dry using a nitrogen gun.

3.7. Data Analysis

1. Open raw data of a typical STR profile in BaseFinder to obtain the color separation matrix parameters for color cross-talk correction (see Note 30).

2. Input the color separation matrices into MegaBACE™ Fragment Profiler 1.2. Create appropriate peak filters for the STR typing kits, the internal lane standard, and bin sets.

3. Convert the data files recorded by the scanner to binary format and append with proper header information by the homemade conversion program.

4. Import the preprocessed data files into the MegaBACE™. Sequentially perform baseline fitting, color cross-talk correction, and allele calling in the program.

4. Notes

1. A detailed protocol of glass microfabrication can be found in Ref. (14). Briefly, the microchannel structure is first patterned on a 100-mm-diameter borofloat glass wafer (Schott, Yonkers, NY, USA) coated with a 2,000-Å film of silicon using standard photolithographic procedure. The photopolymer representing the microchannel pattern is removed in the development process, exposing the silicon film, which is removed by reactive ion etching with SF_6 to expose the glass surface. Formation of the glass microchannels is accomplished by isotropic wet chemical

etching of the exposed glass in HF, resulting in a channel depth of 40 µm. The fluidic access holes are diamond drilled using a CNC mill. The wafer is then cleaned, and the remaining silicon film is removed completely by reactive ion etching. The wafer is thermal bonded with a blank wafer to form enclosed channel structures.

2. The design and the operation of the McCAEPs can be found in Ref. (15). Briefly, the instrument contains a 488-nm diode laser, a confocal optical system with a rotary objective for detecting four different fluorescence signals, pneumatics for control of on-chip microvalves, four PCR temperature control systems, and four high-voltage power supplies for electrophoresis. In the optical system shown in Fig. 4c, the laser beam is reflected by two dielectric mirrors, passed through a dichroic beam splitter, and directed up through a hollow shaft of a stepper motor. The beam is displaced 7 mm by a rhomb, and focused into the channels of the microchip. Fluorescence collected by the objective is reflected by the dichroic into a confocal assembly, where the light is focused on a pinhole with an achromat lens, and directed into a 4-color PMT detector by a collimating lens.

3. The use of these continuous reservoirs defined by the PDMS reservoirs is to reduce the total number of electrodes and to provide sufficient buffer capacity for electrophoresis. Since the phosphodiester backbone of DNA is susceptible to protonation and subsequent mobility shift, it is essential to buffer protons and hydroxyls which are electrolytically generated in the wells through the entire analysis to achieve stable and reproducible analysis. The effect of electrophoresis on the chemical composition in the cathode and anode buffer wells and its significance to the reproducibility of electrophoresis was discussed in detail in Ref. (16).

4. A high-pressure filling station is used to load the viscous separation matrix into the microchannels and to wash the chip thoroughly with distilled DI water. The design of the setup is illustrated in Ref. (17). As an alternative, manual loading can be achieved using a syringe mounted with a pipette tip. The opening of the tip should be cut to tightly fit into the wells of the microchip for applying pressure. Any air bubbles in the syringe should be expelled out to avoid possible solution spilling out. Wear safety glasses or a facial mask when using the syringe.

5. We synthesized the polyDuramide in lyophilized form in house. See Ref. (18) for a more detailed protocol. The prepared coating solution is best to store at ~4°C (up to 3 months).

6. Monomeric acrylamide is neurotoxic and can be absorbed through skin or inhaled (powder). Appropriate precautions

should be taken to avoid skin contact when handling the bis-acrylamide solution.

7. The photoiniator, riboflavin, is sensitive to room light. Prepare the solution in a dark room with a red light illumination.

8. Be careful with TEMED since it has a strong irritating smell.

9. The LPA is synthesized in-house. The protocol can be found in Ref. (19).

10. The UV exposure system should be located in a dark room with a red lamp for illumination. Protect eyes and skin from exposure to UV light. Carefully read and follow the manual of the mercury arc lamp provided by the manufacturer to avoid any damage to the lamp or bulb explosion.

11. STR typing kits, PowerPlex® 16, PowerPlex® Y, and AmpFlSTR® Profiler Plus®, have been successfully tested on the μCAE microsystem (see Refs. (20) and (21)). To make these STR kits compatible with on-chip sample and inline injection, the unlabeled primers in the primer pairs need to be replaced with the biotin-labeled. For proof of concept, a 9-plex autosomal STR typing system that contains biotin-labeled primers was developed by our group based on the primer sequences and fluorescence dye labeling scheme used in the PowerPlex® 16.

12. The biotin-modified sizing standard was constructed in-house by mixing a series of purified PCR amplicons with different fragment lengths (60, 80, 95, 120, 140, 160, 172, 250, 275, and 350 bp). These fragments are amplified from pUC19 with primer pairs such that one is labeled with ROX and the other labeled with biotin. This sizing standard can be cocaptured and separated with PCR products for allele size calibration.

13. Store formamide frozen at –20°C. Minimize the number of freeze-thaw cycles as much as possible since the quality of the formulation may decrease if exposed to air through frequent sampling.

14. Download BaseFinder from http://bioinfo.unc.edu/glabsoftware/ BaseFinder/BaseFinder.html. It is only available for Mac OS X (see Ref. (22)).

15. The conversion program converts data collected by the detection platform to a form that can be read by the MeagBACE™ Fragment Profiler software. The detailed information of the conversion program can be found in Ref. (23).

16. We found when coating a brand new microchip with polyDuramide, the first run usually produces a poor separation resolution. It is better to do a dummy run after the first coating, and then repeat the normal coating procedure. Start the experiments from the second time.

17. The purpose of the Gel Slick® coating is to make the chip surface, especially that around the sample wells, hydrophobic to keep each sample drop within its respective sample well and avoid mix-up.

18. Gradually turn up the gas flow to avoid sample spill.

19. Streptavidin acrylamide is prone to generating bubbles with excess mixing. Slowly pipette water in the tube 2–3 times to dissolve the powder; do not vortex the tube.

20. The final concentrations of the components in the photopolymerization buffer are: 5% bis-acrylamide (19:1), 4 M Urea, 1× TTE, 2 µg/µL streptavidin, 0.0006% riboflavin, and 0.125% TEMED. The prepared solution can be stored at 4°C in dark for up to 4 weeks without noticeable degradation.

21. Since the sizes of the exposure windows and the microchannels are >200 µm, visually performed manual alignment is accurate enough for the UV exposure purpose.

22. The purpose of adding a drop of LPA gel in each well is to stop hydrodynamic flow in the channels during UV exposure. Other methods, such as adding an equal amount of solution in each well, or sealing the wells with a tape, may also work, but provide less consistency in performance.

23. To improve sensitivity, the amount of samples added into the loading cocktail can be increased to 3.5 µL. Remember to reduce the DI water volume accordingly.

24. If desired, the PDMS reservoirs can be treated in a UV-ozone cleaner for 5 min to enhance the bonding between PDMS and glass.

25. The focus of the objective on the matrix-loaded microchannels is achieved by adjusting the heights of three supporting posts of the stage while monitoring the Raman scattering signal from water in the yellow optical channel (580 nm). Uniform signals should be obtained across all the channels. These signals can also be used to locate the channel positions for the control program to extract data from scanning.

26. To increase the amount of samples injected into the capture gel, the injection time can be increased from 10 min up to 20 min. Do not increase the injection electric field. Too high electric filed results in a lower binding efficiency of biotin-labeled DNA products to the streptavidin capture gel due to the fast movement of DNA.

27. At 67°C, TTE buffer in the waste, cathode, and anode reservoirs will evaporate quickly. Top off the buffer reservoirs with 1× TTE to prevent chip dry out during electrophoresis. The sample reservoirs will not dry out because of formamide.

28. If the capture gel plugs cannot be pushed out by the high-pressure filling station, fill the channels with acetonitrile and let the chip sit for 10 min. Acetonitrile can shrink the polyacrylamide gels and make them easily pushed out.

29. If channels are clogged or very dirty, ashing of the chip in a furnace at 500°C for 3 h can be performed to remove residuals in the microchannels.

30. The typical STR profile should contain well-resolved peaks for each dye labeling. Select these peaks to define the color cross-talk matrix parameters. Additionally, BaseFinder can also be used to process raw data, including baseline fitting, color cross-talk correction, and baseline smoothing. Processed data can be exported to an ASCII text file, and then plotted using data analysis programs, such as Microsoft Excel and Origin.

Acknowledgments

References

1. Hammond, H.A., Jin, L., Zhong, Y., Caskey, C.T., and Chakraborty, R. (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am J Hum Genet* 55, 175–189.

2. Jobling, M.A., and Gill, P. (2004) Encoded evidence: DNA in forensic analysis. *Nat Rev Genet* 5, 739–751.

3. Andersen, J.F., Greenhalgh, M.J., Butler, H.R., Kilpatrick, S.R., Piercy, R.C., Way, K.A., Myhill, H.S., Wright, J.C., Hallett, R., and Parkin, B.H. (1996) Further validation of a multiplex str system for use in routine forensic identity testing. *Forensic Sci Intl* 78, 47–64.

4. Horsman, K.M., Bienvenue, J.M., Blasier, K.R., and Landers, J.P. (2007) Forensic DNA analysis on microfluidic devices: A review. *J Forensic Sci* 52, 784–799.

5. Liu, P., and Mathies, R.A. (2009) Integrated microfluidic systems for high-performance genetic analysis. *Trends in Biotechnol* 27, 572–581.

6. Liu, P., Seo, T.S., Beyor, N., Shin, K.J., Scherer, J.R., and Mathies, R.A. (2007) Integrated portable polymerase chain reaction-capillary electrophoresis microsystem for rapid forensic short tandem repeat typing. *Anal Chem* 79, 1881–1889.

7. Hatch, A.V., Herr, A.E., Throckmorton, D.J., Brennan, J.S., and Singh, A.K. (2006) Integrated preconcentration sds-page of proteins in microchips using photopatterned cross-linked polyacrylamide gels. *Anal Chem* 78, 4976–4984.

8. Herr, A.E., Hatch, A.V., Throckmorton, D.J., Tran, H.M., Brennan, J.S., Giannobile, W.V., and Singh, A.K. (2007) Microfluidic immunoassays as rapid saliva-based clinical diagnostics. *Proc Natl Acad Sci USA* 104, 5268–5273.

9. Gong, M.J., Wehmeyer, K.R., Limbach, P.A., Arias, F., and Heineman, W.R. (2006) On-line sample preconcentration using field-amplified stacking injection in microchip capillary electrophoresis. *Anal Chem* 78, 3730–3737.

10. Yang, H., and Chien, R.L. (2001) Sample stacking in laboratory-on-a-chip devices. *J Chromatogr A* 924, 155–163.

11. Long, Z.C., Shen, Z., Wu, D.P., Qin, J.H., and Lin, B.C. (2007) Integrated multilayer microfluidic device with a nanoporous membrane interconnect for online coupling of solid-phase extraction to microchip electrophoresis. *Lab Chip* 7, 1819–1824.

12. Yeung, S.H.I., Liu, P., Del Bueno, N., Greenspoon, S.A., and Mathies, R.A. (2009) Integrated sample cleanup-capillary electrophoresis microchip for high-performance short tandem repeat genetic analysis. *Anal Chem* 81, 210–217.

13. Liu, P., Scherer, J.R., Greenspoon, S.A., Chiesl, T.N., and Mathies, R.A. (2011) Integrated sample cleanup and capillary array electrophoresis microchip for forensic short tandem repeat analysis. *Forensic Sci Intl: Genetics* 5, 484–492.

14. Simpson, P.C., Wolley, A.T., and Mathies, R.A. (1998) Microfabrication technology for the reduction of capillary array electrophoresis chip. *Biomed Microdevices* 1, 7–26.

15. Scherer, J.R., Liu, P., and Mathies, R.A. (2010) Design and operation of a portable scanner for high performance microchip capillary array electrophoresis. *Rev Sci Instrum* 81, 113105.

16. Bello, M.S. (1996) Electrolytic modification of a buffer during a capillary electrophoresis run. *J Chromatogr A* 744, 81–91.

17. Scherer, J.R., Paegel, B.M., Wedemayer, G.J., Emrich, C.A., Lo, J., Medintz, I.L., and Mathies, R.A. (2001) High-pressure gel loader for capillary array electrophoresis microchannel plates. *Biotechniques* 31, 1150–1153.

18. Albarghouthi, M.N., Buchholz, B.A., Huiberts, P.J., Stein, T.M., and Barron, A.E. (2002) Poly-n-hydroxyethylacrylamide (polyduramide (tm)): A novel, hydrophilic, self-coating polymer matrix for DNA sequencing by capillary electrophoresis. *Electrophoresis* 23, 1429–1440.

19. Ruizmartinez, M.C., Berka, J., Belenkii, A., Foret, F., Miller, A.W., and Karger, B.L. (1993) DNA-sequencing by capillary electrophoresis with replaceable linear polyacrylamide and laser-induced fluorescence detection. *Anal Chem* 65, 2851–2858.

20. Yeung, S.H.I., Greenspoon, S.A., McGuckian, A., Crouse, C.A., Emrich, C.A., Ban, J., and Mathies, R.A. (2006) Rapid and high-throughput forensic short tandem repeat typing using a 96-lane microfabricated capillary array electrophoresis microdevice. *J Forensic Sci* 51, 740–747.

21. Greenspoon, S.A., Yeung, S.H.I., Johnson, K.R., Chu, W.K., Rhee, H.N., McGuckian, A.B., Crouse, C.A., Chiesl, T.N., Barron, A.E., Scherer, J.R., Ban, J.D., and Mathies, R.A. (2008) A forensic laboratory tests the berkeley microfabricated capillary array electrophoresis device. *J Forensic Sci* 53, 828–837.

22. Giddings, M.C., Severin, J., Westphall, M., Wu, J.Z., and Smith, L.M. (1998) A software system for data analysis in automated DNA sequencing. *Genome Res* 8, 644–665.

23. Wedemayer, G.J., Medintz, I., Wong, W.W., Scherer, J.R., and Mathies, R.A. (2001) Conversion of capillary electrophoresis microchip genotyping data for analysis with genetic profiler (tm) software. *Biotechniques* 30, 122–128.

Chapter 25

Microchip Capillary Electrophoresis Protocol to Evaluate Quality and Quantity of mtDNA Amplified Fragments for DNA Sequencing in Forensic Genetics

Coro Fernández and Antonio Alonso

Abstract

Here, we describe a microcapillary electrophoresis technique with application to the quantitative analysis of mtDNA hypervariable regions HVR1, HVR2, and HVR3 PCR amplicons previous to sequence analysis, which yields several important advantages compared to traditional separation and detection methods. Based on laser-induced fluorescence (LIF) detection, and performed in a microchip, this analysis system enables the handling of very small volumes via microchannels etched in the chip. Moreover it is faster than traditional methods; chip priming and sample loading are the only manual interventions, as the rest of the process is fully automated by software control: injection, electrophoretic separation, detection of the fluorescent signal, and calculation of both quantity and size. MtDNA amplicons are separated in microchannels with an effective length of 15 mm and detected by means of the fluorescence displayed by an intercalated dye. A software records the fluorescence and entails the data into size and concentration through the use of two internal standards and an external ladder of 11 fragments. The effectiveness of this procedure has been illustrated with a validation experiment carried out in our laboratory, in order to assess the detection limit of mtDNA sequencing by determining the minimal amount of PCR amplicon needed to edit a reproducible and high quality mtDNA sequence from complementary sequence data obtained using forward and reverse primers.

Key words: Forensic samples, Hypervariable regions, Microchip capillary electrophoresis, Mitochondrial DNA, Quantification, Validation studies

1. Introduction

The sequence analysis of mitochondrial DNA (mtDNA) hypervariable regions I, II, and III (HVR1, HVR2, and HVR3) has been proved to be a useful tool in forensic laboratories for ancient DNA specimens or highly degraded samples which fail to obtain reliable results with nuclear DNA typing because of low DNA

quantity present in the sample or no amount at all as it is the case of hair shafts (1–4). In addition to these forensic applications mtDNA sequencing is also used in kinship analyses of missing people as well as in Mass Fatality Incidents when the reference person is a maternal relative. The sequence analysis of other mtDNA targets (such as Cytochrome b gene) has also been applied to the identification of nonhuman forensic samples (5). A critical quality control step for mtDNA sequencing is establishing the size, purity, and concentration of PCR amplicons. A reliable quantitative analysis of mtDNA HVR1, HVR2, and HVR3 PCR amplicons allows the standardization of the amplicon input in the sequencing reactions, improving the overall quality of sequence data as well as the assessment of the detection limit in different mtDNA sequencing protocols (6).

Traditional methods include agarose or acrylamide gels for the analysis of PCR amplified mtDNA: samples are run in a slab gel together with predetermined standards of known concentration and molecular mass, and separated according to their size. The analysis of the DNA fragments can be achieved in a simple way by a Polaroid photograph taken over a UV light source from which the user calculates size and quantity by matching the band intensity to the standards or by means of more accurate imaging software which interfaces with scanners or digital cameras. This widely used method, it is time-consuming, dangerous materials need to be handled (bromure ethidium), and the accuracy of this type of analysis is minimal.

More recently electrophoresis of mtDNA PCR amplicons has been improved by means of a microchip capillary electrophoresis (MCE) technique, in combination with a laser-induced fluorescence (LIF) DNA detection method (6, 7). Samples are loaded in a small chip together with a sieving polymer and fluorescence dye. The loaded chip is placed into an analyzer where 16 platinum electrodes fit into the wells, becoming the chip an integrated electrical circuit. MtDNA amplicons are separated in microchannels with an effective length of 15 mm and detected by means of the fluorescence displayed by the intercalated dye. A software records the fluorescence and entails the data into size and concentration through the use of two internal standards and an external ladder of 11 fragments. MCE offers several advantages over conventional slab methods: reduction in sample and reagent consumption, substantially increased high-throughput and high-speed in quantitative and qualitative analysis of PCR amplicons; moreover, it may perform several DNA analysis tasks, becoming a fully integrated and automated molecular analysis system.

We have already described the application of this commercially available MCE device in different genetic identification studies performed with mtDNA targets, including the haplotype analysis

of HVR1 and HVR2 and the study of interspecies diversity of cytochrome *b* (Cyt *b*) and 16S ribosomal RNA (16S rRNA) mito-chondrial genes in forensic and ancient DNA samples. The method has been proved to be a fast and sensitive detection method of length heteroplasmy in cytosine stretches produced by 16 189 T/C transitions in HVR1 and by 309.1 and 309.2C-insertions in HVR2 (6).

Here we present a complete description of the procedure used for PCR amplification, MCE setup, DNA electrophoresis, and software data analysis with application to the quantitative analysis of mtDNA hypervariable regions HVR1, HVR2, and HVR3 PCR amplicons previous to sequence analysis. The performance of this procedure is illustrated with a validation experiment to assess the detection limit of mtDNA sequencing by determining the minimal amount of PCR amplicon necessary to edit a reproducible and high quality mtDNA sequence from complementary sequence data obtained using forward and reverse primers.

2. Materials

2.1. Specimens

Microchip electrophoresis can be applied to estimate both the size (ranging ± 5% 100–500 pb) and the quantity (ranging 0.1–50 ng/µL) of mitochondrial DNA amplicons obtained from amplifying human DNA extracts (saliva, blood, hair shafts, tissues) and nonhuman DNA of forensic casework interest. Two regions within the control region or displacement loop (D-loop) found in mtDNA are normally examined by PCR amplification prior to sequencing. These regions are known as hypervariable region I (HVR1) and hypervariable region II (HVR2). Less frequently, a third region called hypervariable region III (HVR3) is also amplified.

Standard Reference Material® 2392-I (8) which is a DNA extracted from cell culture line HL-60 at a nominal concentration of 1.4 ng/µL and DNA from cell culture line 9947A at a nominal concentration of 0.1 ng/µL were used for the sensitivity validation experiments.

2.2. PCR Reagents

1. 10× PCR Buffer, 500 mM KCl, 100 mM Tris-HCl, pH 8.3, 15 mM MgCl2, and 0.01% (w/v) gelatin (Applied Biosystems).
2. GeneAmp® dNTP Blend, 10 mM (Applied Biosystems).
3. AmpliTaq Gold DNA polymerase (250 Units) (Applied Biosystems).
4. Autoclaved Bovine Serum Albumin (BSA) (0.05%).
5. Sterile water.

Table 1
Primers pairs (forward and reverse) employed for the amplification of the three hypervariable regions (HVR1, HVR2, and HVR3)

Regions	Primers
HVR1	FORWARD: L15997 5′-CACCATTAGCACCCAAAGCT-3′ REVERSE: H16395 5′-CACGGAGGATGGTGGTCAAG-3′
HVR2	FORWARD: L048 5′-CTCACGGGAGCTCTCCATGC-3′ REVERSE: H408 5′-CTGTTAAAAGTGCATACCGCCA-3′
HVR3	FORWARD: L369 5′-AAACCCCAAAAACAAAGAAC-3′ REVERSE: H600 5′-AACATTTTCAGTGTATTGCT-3′

6. HVR1(L15597-H16395), HVR2 (L048-H408), and HVR3 (L369-H600) PCR primers (100 pmol/μL) (see PCR primer sequences in Table 1).

7. Positive amplification control DNA: genomic DNA culture line 9947A (0.1 ng/μL) (Applied Biosystems).

8. Negative amplification control: Sterile water.

2.3. PCR Equipment

1. Thermal cycler GeneAmp® PCR System 9700 (Applied Biosystems).

2. Laminar flow cabinet.

2.4. MCE Kit Components (see Note 1)

1. DNA chips (Fig. 1a).

2. DNA ladder (11 fragments in the range of 25–1,000 bp).

3. DNA markers (an upper marker of 1,500 bp and a lower marker of 15 bp).

4. DNA dye concentrate.

5. DNA gel matrix.

2.5. MCE Equipment

1. Laminar flow cabinet.

2. Agilent 2100 Analyzer (Agilent) (Fig. 1d, e).

3. Chip priming station (Agilent) (Fig. 1b).

4. Vortex mixer (IKA) (Fig. 1c).

2.6. Software

2100 Expert Software (Agilent)
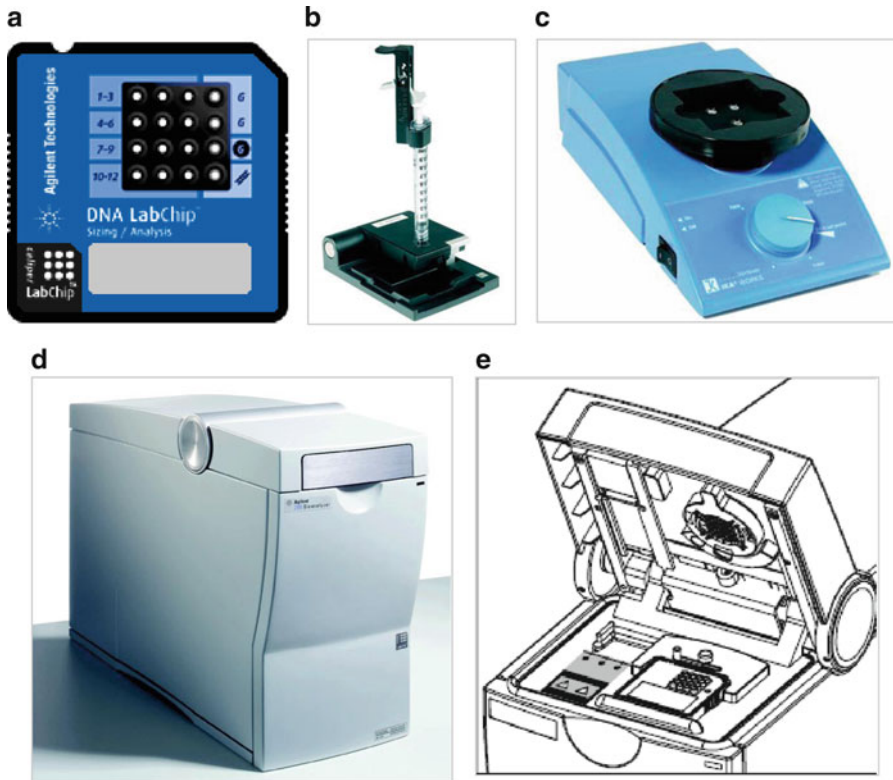
Fig. 1. DNA Chip and MCE equipment. (**a**) DNA chip. (**b**) Chip priming station. (**c**) IKA chip vortex mixer. (**d**) Agilent 2100 Analyzer. (**e**) internal view of Agilent 2100 Analyzer with the DNA chip placed inside.

3. Methods

3.1. PCR Amplification of HVR1, HVR2, and HVR3 from HL60 and 9947A DNA Cell Lines

1. Sample preparation: In both cases, HL60 and 9947A DNA cell lines, dilution series for a PCR input of 100, 10, 1, 0.1, and 0.01 pg of nuclear DNA, were prepared.

2. PCR setup: All PCR reactions were carried out in 50 μl volume reaction, each reaction contained 13.5–20.5 μL of sterile water, 20 μL of BSA, 5 μl of 10× PCR Buffer, 2 μL of dNTPs, 0.5 μL of AmpliTaq Gold DNA polymerase, 0.5 μL of forward primer, 0.5 μL of reverse primer, and 1–8 μL of DNA extract.

3. Calculation of final PCR master mix volume was obtained according to Table 2. N is total number of reactions, including number of samples, positive amplification control, negative amplification control, and an extra reaction to compensate for pipetting errors. The volume of sterile water ranges from 13.5 to 20.5 μL depending on the μl of DNA extract used in the reaction. In particular, for the sensitivity assay 1 μL of DNA extract was used, thereby 20.5 μL of sterile water was added.

Table 2
Calculation of the final PCR master mix volume

Components	Volume
Sterile water	$N \times 13.5$–20.5 μl[a]
BSA	$N \times 20$ μl
PCR Buffer 10X	$N \times 5$ μl
dNTPs	$N \times 2$ μl
Taq GOLD	$N \times 0.5$ μl
Forward primer	$N \times 0.5$ μl
Reverse primer	$N \times 0.5$ μl

N is total number of reactions
[a]The water volume depends on the μl of DNA extract added

4. Use primers sets as follows: HVR1(L15597-H16395), HVR2 (L048-H408), and HVR3 (L369-H600) (Table 1).

5. PCR amplification: it was carried out using a Thermal cycler GeneAmp® PCR System 9700 with the following Thermal cycling conditions: 10 min at 95°C followed by 36 cycles of 95°C for 10 s; 60°C for 30 s, 72°C for 30 s and a final extension of 10 min at 72°C.

3.2. Preparing the Gel-Dye Mix

The gel-dye mix is obtained by mixing DNA dye concentrate and DNA gel matrix in the volumes indicated below. Once prepared the mix should be used within a month.

1. After vortexing DNA dye concentrate for 10 s, pipette 25 μL of it into a 500 μL DNA gel matrix vial. Vortex again for 10 s. Make sure gel and dye are well mixed.

2. Transfer the gel-dye mix into a spin filter, place it in a micro-centrifuge and spin for 15 min at room temperature at $2,240 \times g \pm 20\%$ (for Eppendorf microcentrifuge, this corresponds to 6,000 rpm).

3. Discard the filter following disposing waste material regulations.

3.3. Loading the Gel-Dye Mix

1. Use one microchip each time. Open the chip priming station (Fig. 1b) by pressing the lock of the latch, place the chip on the platform.

2. Pipette 9.0 μL of the gel-dye mix at the bottom of the well marked with a G circled in black. (see Note 2) (Fig. 1a)
Allow the gel to fill the microchannels by closing the Priming Station and pressing the plunger of the syringe down until it is held by the clip.

3. After 60 s exactly release the plunger with the clip release mechanism. It is necessary to check that the plunger moves back at least to the 0.3 mL mark. Wait for 5 s more and then before opening the chip priming station, slowly pull back the plunger to the 1 mL position.

4. Finally, pipette 9.0 µL of the gel-dye mix in each of the two remaining wells marked with G.

3.4. Loading the Marker, the Ladder, and the Samples

1. Pipette 5 µL of DNA marker into each of the wells, except for those which have been loaded with the gel-dye mix.

2. Add 1 µL of the DNA ladder in the well marked with the ladder symbol.

3. Place 1 µL of sample in each of the 12 sample wells. Do not leave any well empty (see Note 3).

4. Place the chip in the IKA vortex mixer and vortex the chip for 60 s at 2,400 rpm (see Note 4).

3.5. Running the Chip in the Bioanalyzer

1. Switch on the Bioanalyzer.

2. Click on the 2100 Expert software icon.

3. Open the lid of the Bioanalyzer and place the microchip properly (Fig. 1e) Close the lid carefully (see Note 5) and choose the appropriate assay: dsDNA/DNA 1000 series.

4. Start the chip running by clicking on the Start button (see Note 6).

5. Once the electrophoresis starts running, in order to enter the sample information select the Data icon, and then select the Chip Summary tab.

6. Complete the sample name table by hand or by importing data.

7. During the electrophoresis raw data can be visualized in the *Instrument* context.

8. After the assay is completed, the used chip should be removed immediately and disposed according to good laboratory practices. Electrodes should be cleaned (see Notes 7 and 8).

3.6. Automatic Data Analysis

Through the use of a ladder of known size, 2100 Expert Software establishes a standard curve by plotting the size of the ladder versus migration time, using a point to point fit. From the standard curve, the software determines the DNA size in base pairs (bp) by means of their migration times. A new standard curve is created in each new assay.

Two internal standards of known sizes (lower marker and upper marker) are added to both the sample and the ladder wells. They are used to align the ladder data with data from sample wells allowing correction for small drifts that may occur during electrophoresis.

In order to quantify DNA fragments, the area under the upper marker peak is compared with the sample peaks areas, as the concentration of this marker is known, concentration for each sample can be calculated.

3.7. Checking Assay Results

1. Select the gel or electropherogram tab in the Data context.

2. In the ladder well window check:
   - If there are 11 peaks displayed for DNA ladder ranging from 25 to 1,000 pb and bracketing them identify the presence of both markers (lower marker 15 pb and upper marker 1,500 pb).
   - If all peaks are well resolved.
   - The presence of a flat baseline.

3. In the sample well window check:
   - That all the peaks appear between the upper and the lower marker and they are well resolved.
   - The presence of a single band or peak electrophoretic pattern. In case of a multiband electrophoretic pattern, this will indicate samples with length heteroplasmy (see Note 9).
   - The presence of a flat baseline.

3.8. Results Evaluation

Once the assay is finished and data checked, a report can be set up containing different information selected by the user: Run summary including the gel photograph, the electropherograms of each sample, concentration, and size of every fragment.

Evaluation and acceptance of results rely on the following items: sizes of amplicons should be the one expected for each of the regions amplified (HVR1: 450–472 bp, HVR2: 403–421 bp, HVR3: 327–332 bp), concentration should be equal or superior to 1 ng/μL (as it has been proved in the sensitivity test performed, described below) in order to obtain a reliable sequence result. Samples fulfilling these requirements will be purified and sequenced.

3.9. Application of Microchip Capillary Electrophoresis to Sensitivity Validation Studies

According to quality forensic DNA standards (9–11) each laboratory should perform internal validation studies to assess the sensitivity levels and stochastic effects of PCR and CE detection protocols. Moreover they should quantify the amount of human DNA in forensic samples. However for mtDNA forensic typing there is not a certified human mtDNA quantification standard, not even a standardized PCR real time protocol for mtDNA quantification of forensic samples.

With those limitations we set out the following approach to carry out sensitivity validation studies for accreditation of our mtDNA laboratory under ISO 17025 standard: We carried out mtDNA analysis of serial dilutions from two DNA standards: HL60 and 9947A Cell Lines (both quantified for Nuclear DNA) by PCR followed by quantification of mtDNA amplicons using MCE
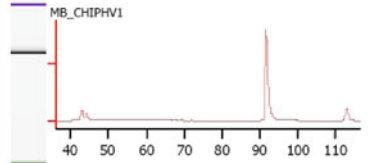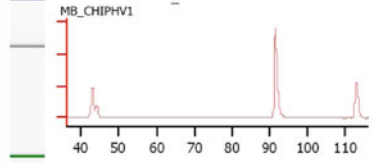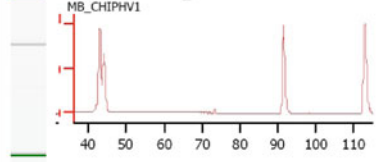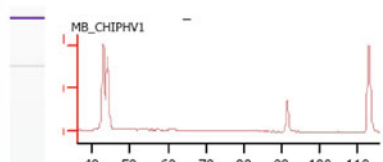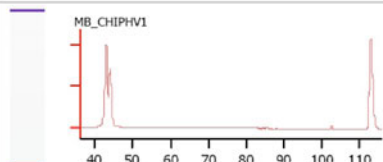
Nuclear DNA PCR Input (pg)	HL60 Dilution Series	HVR1 Amplicon concentration (ng/µl)
100		19.85
10		6.52
1		2.27
0.1		0.83
0.01		undetected

Fig. 2. Gel views, electropherograms, and quantification data of mtDNA HVR1 PCR amplicons using 100, 10, 1, 0.1, and 0.01 pg of nuclear DNA from HL60 cell line as PCR template. The DNA input of 0.1 pg was the last detectable DNA dilution yielding an amplicon concentration of 0.83 ng/µl.

trying to determine the minimal amount of nuclear DNA which entails enough concentration of mtDNA amplicons for a reliable sequence analysis of the three regions HVR1, HVR2, and HVR3.

Figure 2 shows the results of mtDNA (region HVR1) amplicon quantification for cell line HL60 (similar results were obtained for HV2 and HV3 regions, data not shown). On the left side the DNA input in the PCR reaction in pg and on the right side the concentration of amplicon in ng/µL are displayed. As it can be seen the last dilution that resulted in a detectable amount of HVR1 amplicons was 0.1 pg which corresponded with an amplicon

Nuclear DNA PCR Input (pg)	9947A Dilution Series	HVR1 Amplicon concentration (ng/µl)
100	MB_CHIPHV1	19.96
10	MB_CHIPHV1	15.09
1	MB_CHIPHV1	9.49
0.1	MB_CHIPHV1	1.87
0.01	MB_CHIPHV1	undetected
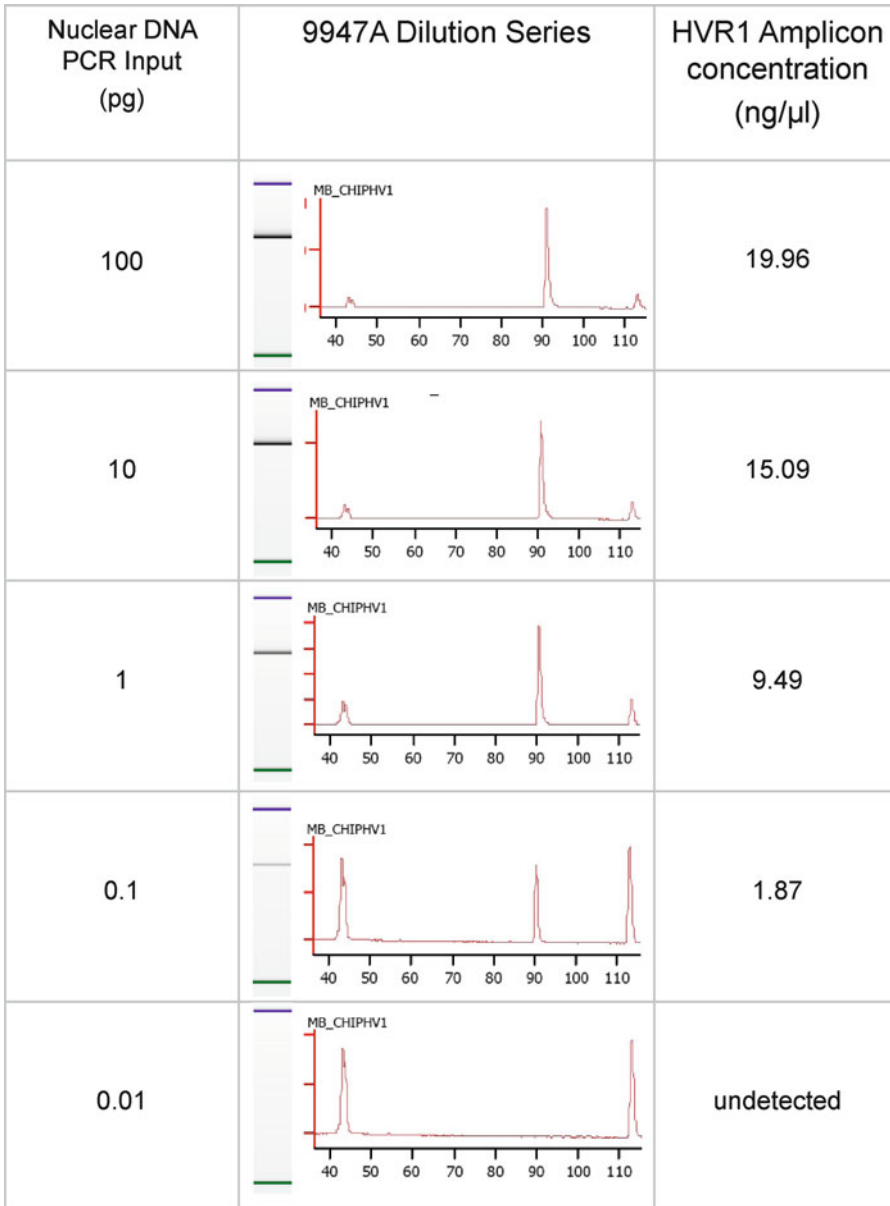
Fig. 3. Gel views, electropherograms, and quantification data of mtDNA HVR1 PCR amplicons using 100, 10, 1, 0.1, and 0.01 pg of nuclear DNA from 9947A cell line as PCR template. The DNA input of 0.1 pg was the last detectable DNA dilution yielding an amplicon concentration of 1.87 ng/µl.

concentration of 0.83 ng/µL. High quality sequence electropherograms were obtained with a DNA input of 10 and 1 pg while in the case of 0.1 pg sequence data were only obtained from one primer (data not shown).

Similar results were obtained for 9947A DNA dilution series (Fig. 3) in this case the DNA input of 0.1 pg was also the last detectable

DNA dilution yielding an amplicon concentration of 1.87 ng/μL. In this case, high quality sequence electropherograms were obtained with a DNA input of even 0.1 pg (data not shown).

To sum up, reliable and reproducible sequence results were obtained for HVR1, HVR2, and HVR3 regions when concentration of mtDNA amplicons was higher than 1 ng/μL.

This mtDNA amplicon concentration threshold was reached even with a nuclear DNA input of:

- 0.1 pg in the case of 9947A DNA cell line.
- 1 pg in the case of HL60 DNA cell line.

(See Note 10 for correspondence between number of genomic mtDNA copies of 9947A DNA cell line and pg of genomic ADN).

4. Notes

1. Stand all reagents and samples at room temperature for 30 min before use. Dye and dye mixtures should be protected from light in order to avoid their decomposition and therefore reducing the signal intensity. Dye should be treated as a potential mutagen as it contains a compound DMSO Methyl Sulfoxide C_2H_6OS which is known to facilitate the entry of organic molecules into tissues. It is necessary to wear lab coat, hand and eye protection when handling and preparing reagents. All waste generated should be treated following good laboratory practices. After use store all reagents at 4°C.

2. When dispensing the gel-dye mix the pipette tip must be inserted to the bottom of the well to prevent air bubble formation otherwise poor results could be achieved.

3. When filled the microchip becomes an integrated electrical circuit so it is necessary to use every well, in case there are not enough samples to complete them dispense 1 μL of deionized water, otherwise the chip will not run properly.

4. Vortex exactly at 2,400 rpm, higher speed could lead to liquid spilling. In order to avoid reagents evaporation, once prepared microchip should be used within 5 min.

5. Pushing the lid carelessly may damage the electrodes and dropping the lid may cause liquid spills resulting in bad results.

6. Avoid touching the Agilent 2100 bioanalyzer during analysis and do not place it on a vibrating surface.

7. Remove the chip immediately after the run has finished and clean the electrodes using the electrode cleaner provided with each new kit in order to avoid contamination of the electrodes. Fill the electrode cleaner with 350 μL deionized analysis-grade

water. Place it on the Bioanalyzer, close the lid, and leave it for about 15–20 s.

8. When performing different microchips in a row, use the cleaning procedure between different assays and wait for 10 s before starting the next run, so as to allow the water on the electrodes to evaporate. Otherwise poor results could be obtained.

9. Mitochondrial DNA length heteroplasmy results from the presence of multiple populations of mtDNA molecules with various lengths within an individual that differ, usually, in the number of cytosines on polycitosine stretches. The sites where heteroplasmy is observed are particularly position 16189 within HVR1 in which a transition T to C creates a long string of cytosines and it also occurs in the HVR2 C-stretch due to C insertions at nucleotides 309 and 315. A multiple pattern indicates the presence of length heteroplasmic C-streches, allowing the use of alternative sequence primers in advance, located upstream and downstream of the C-homopolymeric tract, to generate readable sequences.

10. In order to determine the minimum number of mitochondrial DNA copies for a reliable mtDNA sequencing analysis we have quantified different 9947A DNA lots using two Taqman assays (12) using a cloned human control region standard for quantification of mtDNA copies, (provided by H. Niederstaetter and W. Parson. Institute of Legal Medicine. University of Innsbruck). Different DNA lots from 9947A cell culture line gave around 150–300 copies per 1 pg of Genomic DNA, therefore the detection limit for mtDNA sequencing of 9947A Standard was about 15–30 mtDNA copies (0.1 pg of Genomic DNA).

References

1. 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(2):68–74

2. Holland MM, Parsons TJ (1999) Mitochondrial DNA sequence analysis: Validation and use for forensic casework. Forensic Science Review 11:22–50

3. Carracedo A, Bär W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland M, Tully G, Wilson M (2000) DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int 110(2):79–85

4. Budowle B, Allard MW, Wilson MR, Chakraborty R (2003) Forensics and mitochondrial DNA: applications, debates, and foundations. Annu Rev Genomics Hum Genet 4:119–41

5. Parson W, Pegoraro K, Niederstätter H, Föger M, Steinlechner M (2000) Species identification by means of the cytochrome b gene. Int J Legal Med 114(1–2):23–8

6. Alonso A, Albarran C, Martín P, García P, Capilla J, García O, de la Rua C, Izaguirre N, Pereira F, Pereira L, Amorim A, Sancho M (2006) Usefulness of microchip electrophoresis for the analysis of mitochondrial DNA in forensic and ancient DNA studies. Electrophoresis 27(24):5101–9

7. Kline MC, Vallone PM, Redman JW, Duewer DL, Calloway CD, Butler JM (2005) Mitochondrial DNA typing screens with control region and coding region SNPs. J Forensic Sci 50(2):377–85

8. Levin BC, Holland KA, Hancock DK, Coble M, Parsons TJ, Kienker LJ, Williams DW, Jones M, Richie KL (2003) Comparison of the complete mtDNA genome sequences of human cell lines--HL-60 and GM10742A--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(6): 387–400

9. DNA Advisory Board (2000) Quality assurance standards for forensic DNA testing laboratories July 1998. Forensic Science Communication. http://www.fbi.gov/hq/lab/fsc/backissu/july2000/codis2a.htm. Accessed 10 December 2010

10. Scientific Working Group on DNA Analysis Methods (SWGDAM) (2004) Revised Validation Guidelines. Forensic Science Communications. http://www.fbi.gov/about-us/lab/forensic-science-communications/forensic-science-communications-april-2010. Accessed 10 December 2010

11. Bär W, Brinkmann B, Lincoln P, Mayr W, Rossi U, Budowle B, Eisenberg A, Fourney R, Gill P, Rand S (1992) Editorial: Recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of PCR-based polymorphisms. Forensic Sci Int 55:1–3

12. Alonso A, Martín P, Albarrán C, García P, García O, de Simón LF, García-Hirschfeld J, Sancho M, de La Rúa C, Fernández-Piqueras J (2004) Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies. Forensic Sci Int 139(2–3):141–9

Chapter 26

Microchip Capillary Electrophoresis of Multi-locus VNTR Analysis for Genotyping of *Bacillus Anthracis* and *Yersinia Pestis* in Microbial Forensic Cases

Andrea Ciammaruconi

Abstract

Bacillus anthracis and *Yersinia pestis* are etiological agents of anthrax and plague respectively, and are also considered among the most feared potential bioterrorism agents. These microorganisms show intraspecies genome homogeneity, making strains differentiation difficult, while strains identification and comparison with known genotypes may be crucial for naturally occurring outbreaks vs. bioterrorist events discrimination.

Here an MLVA application for *B. anthracis* and *Y. pestis* strains differentiation is described on Microchip Capillary electrophoresis apparatus. The platform is a candidate for on-site MLVA genotyping of biothreat agents as well as other bacterial pathogens. This microfluidics-based electrophoresis analysis system represents an alternative to the more expensive and demanding capillary electrophoresis methods, and to the less expensive but more time-consuming standard agarose gel approach.

Key words: Lab on a chip, Microfluidic electrophoresis, *Bacillus anthracis*, *Yersinia pestis*, Bacterial genotyping, VNTR

1. Introduction

Bacillus anthracis, a spore forming Gram-positive bacteria, is the etiological agent of anthrax (1, 2), a zoonosis with a worldwide distribution. This bacterium is commonly found in soil, but the disease is transmitted to humans by contact with infected animals or contaminated animal products. Anthrax is still endemic in many countries, Middle East, Africa, North, Central and South America, and other areas of the world (3).

Yersinia pestis, a Gram-negative bacterium, is the causative agent of plague. The bacterium is transmitted by fleas or aerosols,

causing different forms of plague: bubonic, septicemic, or pneumonic (4, 5). Primary pneumonic plague is rapidly progressive and virulent, and, as inhalation anthrax, with mortality rate close to 100% in the absence of a timely treatment. *Y. pestis* and *B. anthracis* are both considered serious threats and potential bioterrorism agents (6) because of their evaluation as bioweapons by Soviet Union and United States laboratories during the past decades. Both agents are classified by the US Centers for Disease Control and Prevention (CDC) in the Bioterrorism Disease Agent List as Category A microorganisms, the most dangerous ones, because of easy dissemination and transmission, high mortality and impact to public health.

B. anthracis and *Y. pestis* both show very low intraspecies genetic diversity (7–9), making very challenging the rapid and accurate differentiation among individual biovars and strains. Nevertheless, finding a way to differentiate the strains by molecular genotyping, remains essential for discrimination between naturally occurring vs. intentional outbreaks. This is the aim of forensic microbiology, as this field is known, and its importance was demonstrated during the 2001 events, and previously by Tokyo (10) and Sverdlovsk (11) incidents. Finally genetic characterization of isolates allows increasing information about worldwide bacterial distribution and epidemiology.

Standard genotyping methods require highly discriminative but heavy, and relatively expensive instruments, such as automated capillary electrophoresis devices, or cheaper, easy to use, but more time consuming and with lower resolution power, such as agarose gels (for a review of bacterial MLVA genotyping see (12)). In the last years, instruments using microfluidic technology for the analysis of biological samples are commercially available. These new miniaturized platforms for quantification and separation of nucleic acid molecules are an evolution of traditional capillary electrophoresis. These systems are also called "Lab on a Chip" and have shown accuracy, precision, and high feasibility along with speed and moderate cost reagents. Here, an application for *B. anthracis* and *Y. pestis* genotyping is described on Agilent 2100 Bioanalyzer (Agilent Technologies). This platform is based on microfluidic technology and allows analyzing 12 DNA samples in 30 min. The device integrates multiple functions onto a single apparatus, so that sample dispensing, separation, detection, and analysis are performed on the same support (a 5×5 cm chip-cast gel). Along with limited weight and size (10 kg, $162 \times 412 \times 290$ mm), the above features make the instrument ideal for field use and other on-site investigations. Finally, Agilent 2100 can also be easily used by low-expertise staff.

2. Materials

1. 1× PCR reaction buffer: 10 mM Tris-HCl, 1.5 mM MgCl2, 50 mM KCl pH 8.3(see Note 1).

2. Taq DNA Polymerase, 5 U/μL.

3. dNTP 10 mM each, pH 7.5, PCR Grade.

4. Nuclease-Free Water.

5. Template DNA prepared as reported in (13) (see Notes 2 and 3).

6. PCR HPSF purified primers from Eurofins MWG Operon (or equivalent).

7. DYAD PTC-220 Thermal Cycler, MJ Research (or equivalent).

8. Agilent 2100 Bioanalyzer (Agilent Technologies).

9. Agilent DNA 1000 Kit (Agilent Technologies).

3. Methods

The 25 VNTR markers amplified for bacterial genotyping were arranged into 12 PCR multiplex reactions either for *B. anthracis* or for *Y. pestis*. This allowed using a single DNA 1000 chip for each strain.

3.1. Genotyping B. anthracis with the 25-loci MLVA

To fit all the 25 loci into a 12-well chip of Agilent 2100 (see Note 4), the *B. anthracis* loci are combined into five triplex, three duplex, and four singleplex (see Table 1) PCR amplifications. The arrangement of different loci in the same multiplex is such as to avoid overlapping of VNTR markers size ranges. For four *B. anthracis* loci (BAMS13, BAMS24, BAMS30, and BAMS34) multiplexing is impossible, because of the large allele size range (see Note 5).

Primer sequences, final primer concentrations, and loci combination are reported in Table 1 and previously in (14). Multiplex PCR reactions are prepared as follows: 5–10 ng of template DNA is amplified in a final volume of 15 μL containing 1× PCR reaction buffer, 0.2 mM dNTPs, the appropriate concentrations of each primer as reported in Table 1, and 1 U of Taq Polymerase (see Note 6) . The 12 multiplex PCR amplifications were performed with a starting denaturation step at 96°C for 3 min, followed by 36 cycles of denaturation at 95°C for 20 s, annealing at 60°C for 30 s, and extension at 72°C for 2 min. The reactions were terminated by a final incubation at 72°C for 5 min.

Table 1
Primer sequences for 25 MLVA loci of *B. anthracis* are reported

	Marker (final primer concentration µM)		Primer sequences
MULTIPLEX 1	CG3 (1.1)	F	TGTCGTTTTACTTCTCTCTCCAATAC
		R	AGTCATTGTTCTGTATAAAGGGCAT
	Bams3s (0.6)	F	GCAGCAACAGAAAACTTCTCTCCAATAACA
		R	TCCTCCCTGAGAACTGCTATCACCTTTAAC
	Bams44 (0.8)	F	GCGAATTAATTGCTCCTCAAAT
		R	GCGAATTAATTGCTCCTCAAAT
MULTIPLEX 2	Bams5 (0.28)	F	GCAGGAAGAACAAAAGAAACTAGAAGAGCA
		R	ATTATTAGCAGGGGCCTCTCCTGCATTACC
	Bams15 (0.6)	F	GTATTTCCCCCAGATACAGTAATCC
		R	GTGTACATGTTGATTCATGCTGTTT
	VrrB2 (0.15)	F	CACAGGCTATTCTTTATCAAACTCATC
		R	CCCAAGGTGAAGATTGTTGTTGA
MULTIPLEX 3	Bams1s (0.6)	F	GTTGAGCATGAGAGGTACCTTGTCCTTTTT
		R	AGTTCAAGCGCCAGAAGGTTATGAGTTATC
	VrrC1 (0.28)	F	GAAGCAAGAAAGTGATGTAGTGGAC
		R	CATTTCCTCAAGTGCTACAGGTTC
MULTIPLEX 4	Bams13 (0.28)	F	AATTGAGAAATTGCTGTACCAAACT
		R	CTAGTGCATTTGACCCTAATCTTGT
MULTIPLEX 5	VrrB1 (0.15)	F	ATAGGTGGTTTTCCGCAAGTT
		R	GATGAGTTTGATAAAGAATAGCCTGTG
	VrrC2 (0.6)	F	CCAGAAGAAGTGGAACCTGTAGCAC
		R	GTCTTTCCATTAATCGCGCTCTATC
	Bams28 (0.6)	F	CTCTGTTGTAACAAAATTTCCGTCT
		R	TATTAAACCAGGCGTTACTTACAGC
MULTIPLEX 6	Bams53 (1.1)	F	GAGGTGTGTTAGGTGGGCTTAC
		R	CATATTTTCACCTTAATTTTGGAAG
	Bams31 (0.6)	F	GCTGTATTTATCGAGCTTCAAAATCT
		R	GGAGTACTGTTTGTTGAATGTTGTTT
MULTIPLEX 7	Bams21 (1.1)	F	TGTAGTGCCAGATTTGTCTTCTGTA
		R	CAAATTTTGAGATGGGAGTTTTACT
	Bams25 (1.1)	F	CCGAATACGTAAGAAATAAATCCAC
		R	TGAAAGATCTTGAAAAACAAGCATT
	VrrA (0.1)	F	CACAACTACCACCGATGGCACA
		R	GCGCGTTTCGTTTGATTCATAC
MULTIPLEX 8	Bams34 (0.4)	F	TGTGCTAAATCATCTTGCTTGG
		R	CAGCAAAATCAATCGAATCAAA
MULTIPLEX 9	Bams24 (0.4)	F	CTTCTACTTCCGTACTTGAAATTGG
		R	CGTCACGTACCATTTAATGTTGTTA

(continued)

Table 1
(continued)

	Marker (final primer concentration μM)		Primer sequences
MULTIPLEX 10	pXO1s (0.6)	F	CAATTTATTAACGATCAGATTAAGTTCA
		R	TCTAGAATTAGTTGCTTCATAATGGC
	Bams22 (0.4)	F	ATCAAAAATTCTTGGCAGACTGA
		R	ACCGTTAATTCACGTTTAGCAGA
	Bams51 (0.4)	F	ATTTCCTGAAGCAGGTTGTGTT
		R	TGCATCTAACAATGCAGAACAA
MULTIPLEX 11	pXO2s (0.28)	F	TCATCCTCTTTTAAGTCTTGGGT
		R	GTGTGATGAACTCCGACGACA
	Bams23 (1.1)	F	CGGTCTGTCTCTATTATTCAGTGGT
		R	CCTGTTGCTCCTAGTGATTTCTTAC
MULTIPLEX 12	Bams30 (0.8)	F	AGCTAATCACCTACAACACCTGGTA
		R	CAGAAAATATTGGACCTACCTTCC

Final μM concentrations for every primer pairs are indicated in brackets

3.2. Genotyping Y. pestis with the 25-loci MLVA

To fit all the 25 loci into 12 chip wells, the ϒ. *pestis* 25 markers are grouped into 11 duplex and one triplex (see Table 2). Although majority of primer sequences are as previously described in (15) eight ϒ. *pestis* markers primer pairs were redesigned to allow multiplexing (see Table 2). Thus, final amplicon sizes are increased to avoid any possible overlap with other different loci into the new multiplexed reactions (see Note 5). Reaction conditions and PCR thermal profile are the same as described for *B. anthracis.*

3.3. Preparation of "Lab on a Chip" Assay

Each reaction is loaded into chip wells prepared according to manufacturer's recommendations (DNA 1000 LabChip Kit) (see Note 7). Each chip contains 16 wells, 12 for the samples (see Note 8), 3 for gel mix (1 labeled black **G** and 2 gray **G**), and 1 for the ladder. Briefly, the black **G** well of DNA Lab Chip support is filled with 9 μL of a gel containing intercalating dye (see Note 9). The DNA Lab Chip is pressurized by a syringe in a priming station device and released after 60 s. Then, two gray **G** wells are filled with 9 μL of gel-dye mix. After gel preparation, each sample well is loaded with 1 μL of PCR reaction and 5 μL of internal marker (containing two MW size standards of 15 and 1,500 bp). Finally 1 μL of DNA ladder is loaded in the ladder well, the chip vortexed at 2,400 rpm for 60 s, and inserted into Agilent 2100 Bioanalyzer. During the run the instrument analyzes sequentially every sample, showing electropherogram, virtual gel image, and table data (see Note 10).

Table 2
Primer sequences for 25 MLVA loci of _Y. pestis_ are reported

	Marker (final primer concentration μM)		Primer sequences
MULTIPLEX 1	Ypms01 (0.2)	F	ACTTCGATGATTATTTTGTGCGTA
		R	TGTTTGGTGCTATTGCCGTA
	Ypms51 (0.2)	F	GGTTTTTACCGATATAAATCCTGAG
		R	GACCAAGAAGTTAAGTTGCTTATCG
MULTIPLEX 2	Ypms04 (0.2)	F	CGCTGTTGAAGTTTTAGTGTAAGAA
		R	AAATGTAACCTGCCAAACGTG
	Ypms62 (0.2)	F	AAACGGTCGTTAACGGAAGA
		R	GCTGAACAGCCCCATAAAAC
MULTIPLEX 3	Ypms05 (0.2)	F	CCTCAGTTCATTGTGTAAAATCTCA
		R	GTATTAGCGAGATCACAGATGAGC
	Ypms07 (0.2)	F	ATACCGCTACGATCAGCCTCTAT
		R	ATTTAATATTGATTTTGGGACTTGC
MULTIPLEX 4	Ypms74 (0.2)	F	CCCCGACTTATATCAAGCACTG
		R	AACTGACGATCTTTTTCACTGAGTT
	Ypms15 (0.2)	F	ACAAGCAAGTTGCGCAGATA
		R	CAAGTTCGCTTTCTGTGTCG
MULTIPLEX 5	Ypms35 (0.2)	F	CTGTTACCGGTCAAAGTGGATATT
		R	AGGCTCTCCTTATCATTATTTGGTC
	Ypms40 (0.2)	F	TCCTGCTGCTGAGTTCATCTT
		R	TCATGTGCAATAGGCGTTGT
MULTIPLEX 6	Ypms20 (0.2)	F	GCCAGAATAATGGCCGTAAA
		R	GTGGTTGTTCTTCACGTTGC
	Ypms38 (0.2)	F	GTGAGGTATAGCTAAACGGTGATGT
		R	CGCCGTAGATTATTTGTCACTTTAT
MULTIPLEX 7	Ypms41 (0.2)	F	GAAGAAAGCCAGCTAATCTGATG
		R	TAATGAATAGCAACGACAACCAATA
	Ypms56 (0.2)	F	GCTCTAACAACGCCGGTAAA
		R	ATGGCATCAACCGACTGACT
MULTIPLEX 8	Ypms44 (0.2)	F	CAATTCCAACAGCTATTAATGCAA
		R	GAATTTTCATAACACGTTCTTCCTG
	Ypms45 (0.2)	F	GCATCGGAGACTGGGTAAAC
		R	TTTCTGAGGATTTATCGGTGTGAT
	Ypms09 (0.2)	F	TTTTTGCCGCACTGAACATA
		R	TCTGATGATTGTTGTGGTCGT
MULTIPLEX 9	Ypms71 (0.2)	F	GCTACTCGAATATGAGTTAGCCAAA
		R	ATTGCCATATTGGATGCTAAAATAA
	Ypms46 (0.2)	F	CAGGTTTTACGTTATTTTCTGAAGG
		R	CAGCATGAAGTATGACGGGTATATTA

(continued)

Table 2
(continued)

	Marker (final primer concentration μM)		Primer sequences
MULTIPLEX 10	Ypms54 (0.2)	F	GTCCACCATTTTCATACTGTCACTT
		R	GCTCTTTGTTCGATTTTATTGAATG
	Ypms70 (0.2)	F	AAACCAACGGTTCATATTGAATAAA
		R	CTTCTTCCGCTATTTTCCTACAGA
MULTIPLEX 11	Ypms69 (0.2)	F	GACGTTGCAACTGCAAAAATAAG
		R	ACTTGTTGTGAAGACCATCACTCT
	Ypms06 (0.2)	F	AATTTTGCTCCCCAAATAGCAT
		R	TTTTCCCCATTAGCGAAATAAGTA
MULTIPLEX 12	Ypms21 (0.2)	F	ACCGGCTTAAAGCAGATTGA
		R	TTCCCTGTACTCGATTGTTGTG
	Ypms73 (0.2)	F	AATACCCTGTGGGTGATAATGAAC
		R	ATCGATTTAGGTACCACCAATTCA

Final μM concentrations for every primer pairs are indicated in brackets

3.4. Data Analysis   Under the described amplification conditions, each expected fragment is amplified (see Note 11), and the software Agilent 2100 Expert version B.02.03.SI307, firmware C.01.055 can identify amplicon size (in Fig. 1a *B. anthracis* triplex electropherogram is shown) (see Note 12). The vast majority of data produced by Agilent 2100 are not accurate. The discrepancies between observed and expected fragment sizes are, presumably, the result of abnormal migration patterns of some repetitive sequences into electrophoretic matrix. The offset and conversion from observed to expected data (and correspondent allele) can be found in comparison Table 1 and Table 2 published in (14). For *Y. pestis* markers (for which repeat units are at least 7 bp long), variability observed is up to two bases (at the shortest repeat units) and, as previously described for *B. anthracis*, discrepancies do not exceed 25% of the repeat unit size.

For chromosomal markers (all the loci, but pXO1 and pXO2 of *B. anthracis*), reproducibility of the observed data allowed correct assignment of each allele for every locus (<25% of unit length). However, for *B. anthracis* plasmid markers (pXO1 and pXO2), since their shorter repeat units and PCR product sizes compared to the chromosomal ones, we observed variability not exceeding one single additional base for each allele in repeated runs. Even with this additional base, correct assignment could be reproducibly done for the plasmid loci. For each individual strain the Agilent assay is completed in about 30–40 min after PCR amplification. For a single strain to be genotyped, this time is significantly lower than agarose gel or automated sequencer fragment separation.
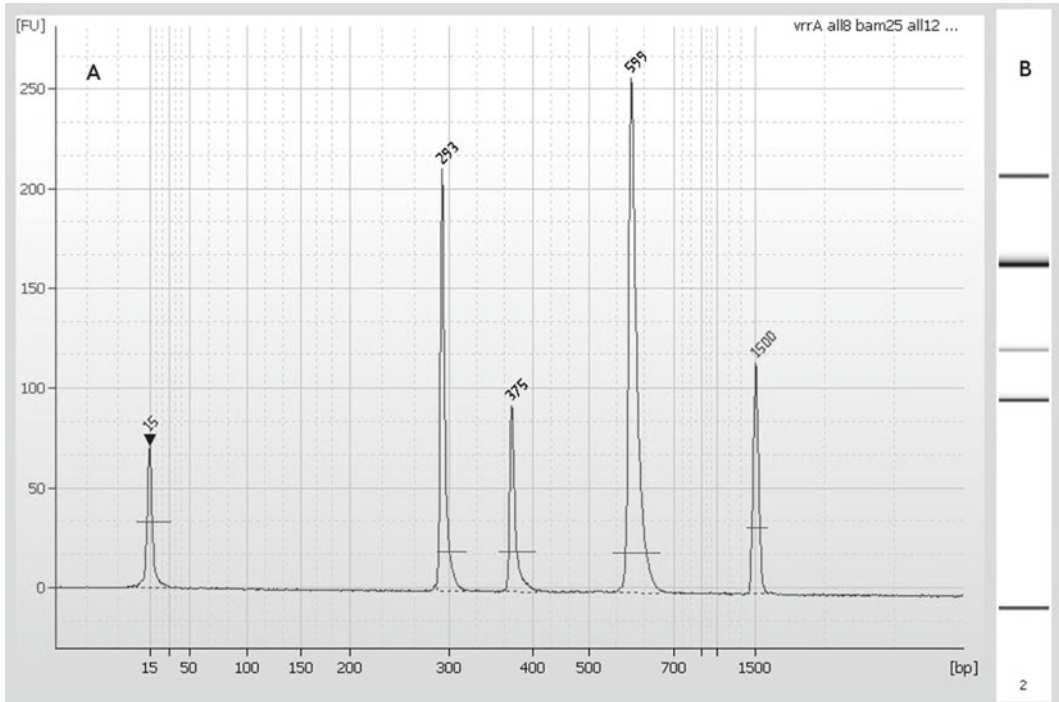
Fig. 1. *B. anthracis* multiplex n. seven electropherogram (**a**) and virtual gel (**b**). For every amplicon, the observed size of markers VrrA, BAMS25, and BAMS21 is shown.

3.5. Comparison of the Microchip Capillary Electrophoresis Method vs. Agarose Gel Electrophoresis Method

1. Sensitivity: Agilent 2100 detects fragments down to 0.1 ng, under routine staining condition, Agarose gel with Ethidium Bromure has a detection limit of 2–5 ng.

2. "Lab on a Chip" allows each individual fragment is automatically sized and quantified.

3. "Lab on a Chip" platforms have better resolution than Agarose gel. It is possible to separate fragments down to 5 bp.

4. "Lab on a Chip" has minimal sample consumption, only 1 µl of material required per analysis.

3.6. Reproducibility

The reproducibility of the "Lab on a Chip" is a critical factor in the validation of the method and its acceptability for bacterial genotyping. In order to check if data were reproducible, we tested interchip and intrachip variability in markers containing the smallest repeat units, for which incorrect allelic calling is more probable. The samples were run in triplicate on the same and on different chips, and size data compared. In general, a low level of interchip/intrachip variability (less than 25% of the repeat unit length) was observed. For example, for *B. anthracis* CG3 marker (Unit Length = 5 bp, so 25% of the Unit Length is 1.25 bp), we observed

168 ± 1 for allele 1 and 173 ± 1 for allele 2. To correctly convert the Agilent DNA fragment size estimates into repeat copy numbers, it was necessary to establish conversion tables (Tables 1 and 2 of (14)) containing 2100 Bioanalyzer fragment sizes as well as actual sizes corresponding to repeats unit numbers.

4. Notes

1. *B. anthracis* and *Υ. pestis* cells must be handled exclusively under Bio Safety Level 3 (BSL3) containment conditions.

2. After extraction, *B. anthracis* DNA templates must be tested for the absence of spores. 30–50 μL of DNA aliquots are plated on blood agar plates and incubated for up to 1 week.

3. During multiplexing, possible overlapping of different loci alleles must be avoided. An interval of 30–40 bp is recommended between the major allele of shorter locus and the minor allele of larger following locus.

4. Care must be taken in PCR primers design. The instrument sizing range is 25–1,000 bp, so larger alleles could not be analyzed.

5. Primer concentrations can be experimentally adjusted to obtain a balanced amplification of different fragments. This optimization is facilitated using the quantification data supplied by the 2100 Bioanalyzer (peak area and fluorescence peak level).

6. Artifact bands (also called "stuttering" bands or "ghost" bands) are sometimes reported when VNTR regions are amplified. These products are minor bands with usually one repeat length smaller than the main bands. Artifact decrease can be observed at 65°C as PCR extension temperature.

7. The maximum salt concentrations allowed in samples are 250 mM for KCl, 15 mM for MgCl2 250 mM NaCl.

8. If samples are less than 12, fill all the remaining wells with buffer or sample replicate. It is not possible to reuse the chip.

9. Chip preparation takes 10–15 min, while the run takes approximately 30 min. The chip must be prepared freshly. The time between preparation and chip run should not exceed 5 min.

10. Store PCR reagents at –20°C.

11. Store Agilent reagents at 4°C.

12. Protect Agilent gel matrix containing intercalating dye from light and store in the dark.

References

1. Turnbull PC (2002) Introduction: anthrax history, disease and ecology. Curr Top Microbiol Immunol 271:1–19.

2. Mock M, Fouet A (2001) Anthrax. Annu Rev Microbiol 55:647–671.

3. Van Ness GB (1971) Ecology of anthrax. Science 172:303–307.

4. Perry RD, Fetherston JD (1997) *Yersinia pestis* – etiologic agent of plague. Clin Microbiol Rev, 10:35–66.

5. Rollins SE, Rollins SM, Ryan ET (2003) *Yersinia pestis* and the plague. Am J Clin Pathol 119 (Suppl):S78–S85.

6. Bossi P, Garin D, Guihot A, Gay F, Crance JM, Debord T, Autran B, Bricaire F (2006) Bioterrorism: management of major biological agents. Cell Mol Life Sci 63:2196–2212.

7. Price LB, Hugh-Jones M, Jackson PJ, Keim P (1999) Genetic diversity in the protective antigen gene of *Bacillus anthracis.* J Bacteriol 181:2358–2362.

8. Jackson PJ, Hill KK, Laker MT, Ticknor LO, Keim P (1999) Genetic comparison of *Bacillus anthracis* and its close relatives using amplified fragment length polymorphism and polymerase chain reaction analysis. J Appl Microbiol 87:263–269.

9. Adair DM, Worsham, PL, Hill KK, Klevytska AM, Jackson PJ, Friedlander AM, Keim P (2000) Diversity in a variable-number tandem repeat from *Yersinia pestis.* J Clin Microbiol 38:1516–1519.

10. Takahashi H, Keim P, Kaufmann AF, Keys C, Smith KL, Taniguchi K, Inouye S, Kurata T (2004) Bacillus anthracis incident, Kameido, Tokyo, 1993. Emerg Infect Dis 10:117–120.

11. Meselson M, Guillemin J, Hugh-Jones M, Langmuir A, Popova I, Shelokov A, Yampolskaya, O (1994) The Sverdlovsk anthrax outbreak of 1979. Science 266:1202–1208.

12. Vergnaud G, Pourcel C (2006) Multiple Locus VNTR (Variable Number of Tandem Repeat) Analysis (MLVA). In: Molecular Identification, Systematics and Population Structure of Prokaryotes. Stackebrandt E. Springer-Verlag, 83–104.

13. Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson PJ, Hugh-Jones ME (2000) Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis.* J Bacteriol 182: 2928–2936.

14. Ciammaruconi A, Grassi S, De Santis R, Faggioni G, Pittiglio V, D'Amelio R, Carattoli A, Cassone A, Vergnaud G, Lista F (2008) Fieldable genotyping of Bacillus anthracis and *Yersinia pestis* based on 25-loci Multi Locus VNTR Analysis. BMC Microbiol. Jan 29, 8:21.

15. Le Fleche P, Hauck Y, Onteniente L, Prieur A, Denoeud F, Ramisse V, Sylvestre P, Benson G, Ramisse F, Vergnaud G (2001) A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis.* BMC Microbiol 1:2.

INDEX