2022

mtDNA Heteroplasmy in Hair Shafts versus Buccal Swabs for Forensic Applications

Sara R. Bodnar
srb0069@mix.wvu.edu

mtDNA Heteroplasmy in Hair Shafts versus Buccal Swabs for Forensic Applications

Sara R. Bodnar

Thesis submitted to the Eberly College of Arts and Sciences at West Virginia University

in partial fulfillment of the requirements for the degree of

Master of Science in
Forensic & Investigative Science

Tina Moroose, M.S., Chair
Arati Iyengar, Ph.D.
Stephanie Young, Ph.D.

Department of Forensic & Investigative Science

Morgantown, West Virginia
2022

Keywords: Mitochondrial DNA, Heteroplasmy, Hair Shaft

ABSTRACT

mtDNA Heteroplasmy in Hair Shafts versus Buccal Swabs for Forensic Applications

Sara R. Bodnar, B.S.

Mitochondrial DNA (mtDNA) analysis has several forensic applications such as criminal investigations, identification of human remains, and missing person investigations. It is also the only type of DNA that is available from certain sample types such as hair shafts. The presence of mtDNA heteroplasmy within and between tissue types can lead to mtDNA sequence differences when comparing samples originating from the same individual. Studies on mtDNA heteroplasmy are increasingly being carried out for their implications in forensic interpretation of mtDNA sequences. Specifically, mtDNA in buccal swabs compared to hair samples from one individual may show differences in sequence due to heteroplasmy, and casework samples compared to reference swabs may exhibit differences that must be correctly interpreted to prevent faulty conclusions made by investigators and scientists alike. Establishing expected rates of heteroplasmy in mtDNA extracted from hair shaft samples and comparison to mtDNA extracted from buccal swab samples will lead to increased confidence in mtDNA interpretation.

The goals of this study were to (1) successfully sequence the entire mtDNA control region from buccal swab samples from 5 volunteers using Sanger sequencing, (2) amplify smaller (<300bp) sections of overlapping regions of the mtDNA control region from 15-20 hair shafts collected from three areas of the scalp using three methods of DNA extraction, and (3) evaluate mtDNA sequences from hair shafts and buccal swabs to identify heteroplasmy within and between samples. Overall, only 20% of the extracted hair samples resulted in half of Hypervariable Region 1 (HV1) being successfully sequenced from either the 5′ or 3′ end. Two out of five participants showed length heteroplasmy in the poly-cytosine region beginning at position 303 within the HV2 region. Point heteroplasmy was observed in one participant at one position in the buccal swab (nucleotide position 16093) as well as at two different positions in a hair sample (nucleotide positions 16258 and 16288) that did not show heteroplasmy in the buccal swab. The heteroplasmy seen in the buccal swab could not be compared to the hair sample as position 16093 did not fall within the successfully sequenced region in the hair. Although only a small subset of hair shafts were successfully sequenced, this study has succeeded in showing that mtDNA heteroplasmy seen in a hair shaft may not be present in buccal swab mtDNA. Further research into rates of heteroplasmy in hair shafts vs. buccal swabs is paramount to bettering the interpretative abilities of forensic scientists working with mtDNA and preventing false exclusions.

Contents

1. Introduction

1.1 Background

Biological evidence has been at the crux of forensic science since the advent of DNA analysis and comparison. Since the first use of DNA fingerprinting by Dr. Alec Jeffreys in 1985, forensic uses of DNA have continued to grow and improve [1]. Typically, DNA is extracted from biological fluids such as blood, saliva, and semen [2]. When nuclear DNA is in low quantities or degraded e.g. in hair shafts, mitochondrial DNA (mtDNA) may be extracted and subsequently tested [3, 4]. The first use of mtDNA in a court case was in 1996 where mtDNA results were entered as evidence against Paul Ware on trial for murder [5]. This case was also the first use of mtDNA extracted from a hair sample [5]. Apart from criminal cases, mtDNA has been utilized in the identification of victims of disaster as well as unidentified remains [3]. Resistance to degradation as well as its abundance in hair shafts, bones, and teeth relative to nuclear DNA makes mtDNA more reliable in terms of providing usable profiles for comparison [2]. While mtDNA provides benefits when faced with certain evidentiary samples, the decreased discriminatory power when compared to nuclear DNA and the extra precautions needed when handling such biological samples has led to only a handful of laboratories across the United States providing any type of mtDNA services for casework.

Attempts have been made to simplify the process of mtDNA extraction and processing through new methods for hair extraction, sequence alignment, and high through-put Sanger sequencing [6, 7]. Commercial kits are also available for the extraction, quantitation, and amplification of mtDNA [8–10]. Research using massively parallel sequencing (MPS) has targeted the entire mtDNA control region to assess heteroplasmy rates in hair versus blood and buccal samples [11]. MPS is much more sensitive and novel than traditional Sanger sequencing with the former able to detect mutant alleles down to ≥5% versus 15-20% using Sanger sequencing [12, 13]. Regardless of the increased sensitivity of MPS, forensic laboratories are simply not equipped with the instrumentation needed for such current methods.

1.2 Nuclear Deoxyribonucleic acid (DNA)

The chromosomes that make up the nuclear human genome and are inherited from both parents confer the traits passed down to each new generation of offspring [2]. Approximately 1% of human DNA differs between individuals, and these differences can be used in identification of source DNA [2]. Non-coding regions of chromosomes are used in the identification of individuals as differences between two people are more abundant [2]. Such differences are known as polymorphisms [2]. Polymorphisms can be divided in two main types: sequence polymorphisms and length polymorphisms [2]. Sequence polymorphisms are differences in the individual nitrogenous base pairs [2]. Length polymorphisms are differences in the number of times a portion of the DNA sequence is repeated [2]. The most common type of repeat is a short tandem repeat (STR), comprised of consecutive repeats of a sequence typically from two to seven base-pairs in

length [2]. STRs are compared in forensic science to determine if two or more samples of DNA are from the same source [2].

Due to the variability in the STR loci used in forensic science, STR analysis is highly discriminatory [2]. Population studies have allowed for the determination of allelic frequencies across various populations. Random match probabilities are calculated using the STR profile of unknown evidence samples to provide statistical weight to the finding of a "match" between an evidentiary sample and a suspect. [2]. Oftentimes, this probability of a randomly selected, unrelated individual from a given population having the same profile as the evidentiary sample can be smaller than one in a trillion, showing the discriminatory power of STR analysis [2]. Such high discriminatory power coupled with the use of multiplexing of the twenty Core CODIS Loci presently targeted has led to the popularity of such analyses [2, 14].

1.3 Mitochondrial DNA (mtDNA)

While each human cell only contains two copies of the entire nuclear DNA genome, mitochondrial DNA (mtDNA) is much more abundant [15]. Within the cell, mitochondria function as the cell's energy source by producing adenosine triphosphate (ATP). Each cell has thousands of mitochondria, each with between two and ten copies of the full mitochondrial genome [3]. Unlike most cellular organelles, mitochondria boast a double-membrane with the inner membrane being the site of oxidative phosphorylation [15]. In mammals, mtDNA is believed to only be inherited from one's mother as the sperm's mitochondria-rich midpiece detaches from the sperm head upon fertilization of an egg [16]. Since mtDNA only reflects DNA passed down from one parent's lineage, nuclear DNA is much more discriminating when attempting to identify an individual through DNA. The evolution of mtDNA down a maternal line can cause differences in the mitochondrial genome and thus create different haplotypes, and groups of similar haplotypes (haplogroups) of individuals sharing a common ancestor with each mutation [17].

Figure 1.1 shows the complete human mitochondrial genome at 16,569 base pairs in length [2]. Unlike the linear chromosomes of the nuclear genome, the mitochondrial genome is comprised of only one circular molecule and has a single non-coding region [2]. The double stranded mitochondrial genome is composed of a heavy strand and a light strand [2]. More guanine residues are present in the heavy strand, leading to this differentiation [2, 3]. Thirty-seven genes are encoded within the mitochondrial genome, including 22 tRNAs, a 23 S rRNA, a 16 S rRNA, and 13 protein coding genes [16]. mtDNA originated from a symbiotic relationship between a eukaryote-precursor and an aerobic prokaryote with the latter eventually becoming what is known today as the mitochondrion [18]. These organisms exchanged genetic material, with the prokaryote eventually experiencing a reduction in its genes as some material was then exported outside of the ancestral eukaryote [15]. Over time, only genes useful to cellular respiration remained within the mitochondria, leading to its small size [15]. The control region is the most variable region of the genome and is a mere 1,210 base pairs long [15]. In the control region, three hypervariable regions (HV1, HV2, and HV3) make up the portions with the highest levels of variability and are the regions most commonly investigated for forensic applications [2].

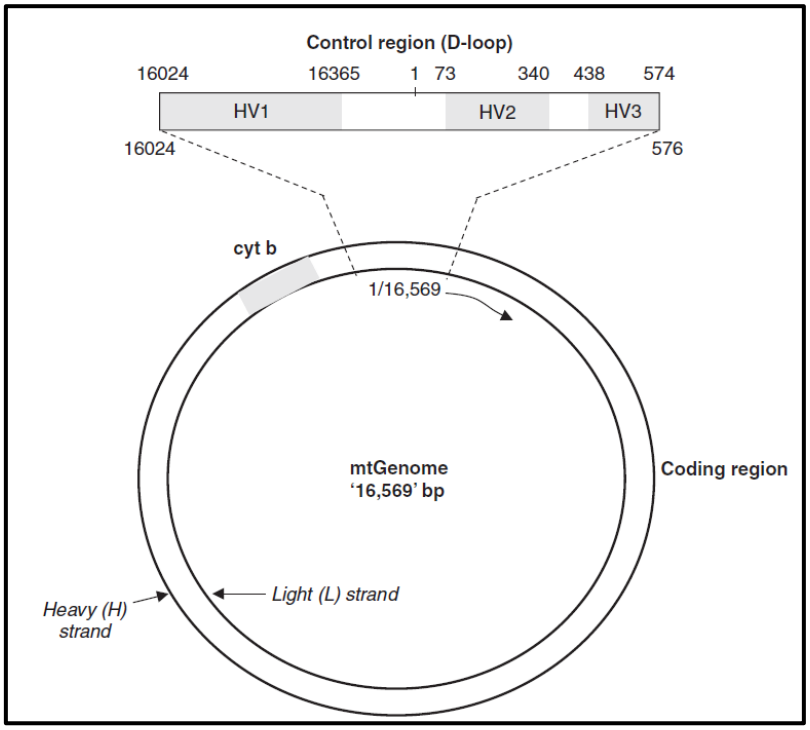

Figure 1.1: Mitochondrial Genome [2]. The circular mitochondrial genome is only 16,569 base pairs in size. The 1,210 base pair control region is the most variable region of the genome and includes three Hypervariable (HV) Regions.

1.3.1 Mitochondrial Haplogroups

Haplogroups for mtDNA are comprised of mtDNA haplotypes within groups of people who genetically share a common maternal ancestor and can be determined through mtDNA sequencing [19]. Over generations, new mutations or polymorphisms arise and are passed down in combination with prior mutations along maternal lineages which can be used to study mtDNA phylogeny [20]. Haplogroups have been studied in populations across nearly all geographic locations with certain haplogroups being most prevalent in specific geographic regions as a consequence of the patterns of human evolution and migration out of Africa (**Figure 1.2)** [17, 21]. The ability to determine ancestry through one's mtDNA haplogroup has led to the development of ancestry tests through companies such as Genebase and Family Tree DNA™ [22, 23]. The mitochondria's important role in health has also led to many studies investigating links between haplogroups and various diseases [24, 25]. Associations between certain haplogroups and diseases such as primary open-angle glaucoma and non-alcoholic fatty liver disease have been reported [24, 25].

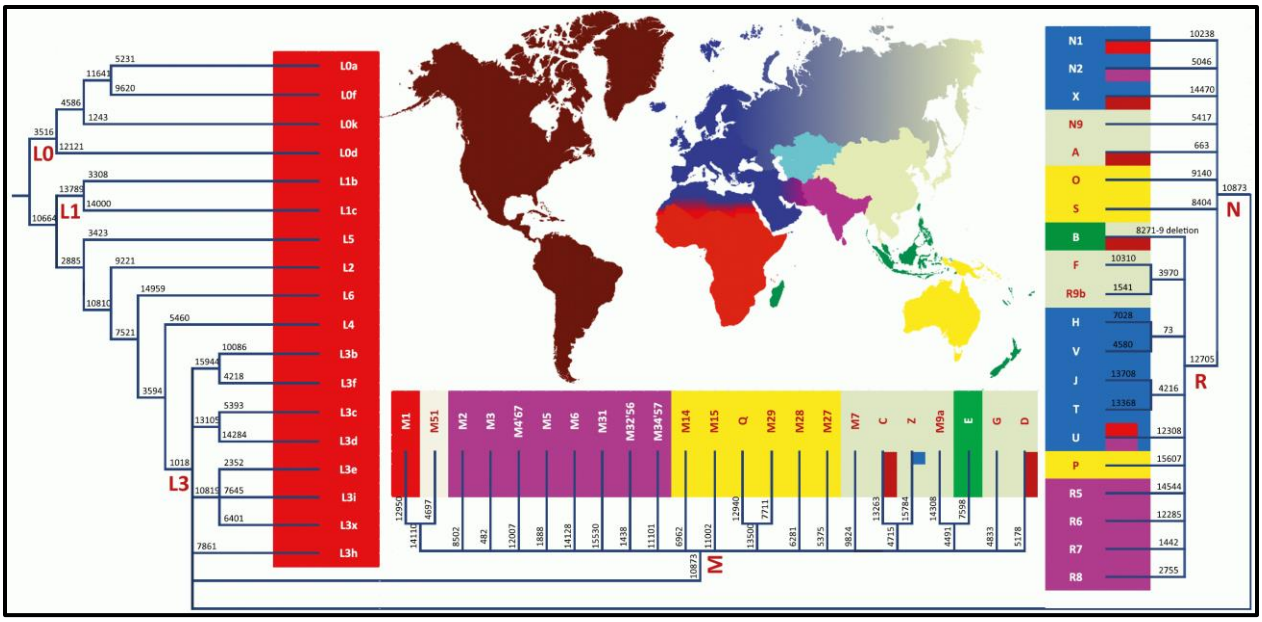

Figure 1.2: Map of mtDNA Haplogroup Phylogeny tree and Geographic Distribution [26]. *The phylogenetic tree begins with the ancestral haplogroup L. Only a single branch-defining marker is provided for each haplogroup.*

1.4 Forensic Uses of mtDNA

In cases of advanced DNA degradation or tissues containing little to no nuclear DNA, mtDNA has been successfully sequenced. The sheer quantity of mtDNA available in each cell makes mtDNA sequencing more sensitive than nuclear DNA, especially when samples with limited DNA are received as evidence. While mtDNA is more abundant than nuclear DNA, the former is not nearly as discriminating in terms of identification [2]. Individuals in the same maternal line cannot be distinguished via the comparison of mtDNA [2]. Conversely, the maternal inheritance of mtDNA allows for the identification of missing persons or mass disasters through mtDNA analysis of suspected matrilineal relatives, even one several generations removed [27]. The Combined DNA Index System (CODIS) has mtDNA databases that can be used in the investigations of missing persons or familial identification [2].

Current mitochondrial testing involves the amplification of the HV1 and HV2 regions using primers across more conserved regions [28]. The FBI utilizes primers L15997 and H16391 to target HV1 and primers L048 and HH408 to target HV2 where L and H refer to the positions on the light and heavy strands, respectively [27]. mtDNA sequences encompassing both HV1 and HV2 from evidence and reference samples are aligned with the revised Cambridge Reference Sequence (rCRS) [29] and then compared to determine whether an exclusion, failure to exclude, or inconclusive result is appropriate [27]. All sequences differences from the rCRS for each sample are recorded during this process[27]. To exclude an evidentiary sample from coming from a known source, there must be at least two nucleotide differences (excluding any length heteroplasmy) between the questioned and known sample[27, 30]. The lack of nucleotide differences at any position within HV1 or HV2 would result in a failure to exclude determination [27, 30].

Comparisons are inconclusive if there is only one nucleotide difference between the samples [27, 30]. While HV1 and HV2 are the minimum requirements for comparison, additional portions of the control region such as HV3 can be sequenced [27, 30].

mtDNA testing is preferred in cases of severe nuclear DNA degradation or when the only evidence is hair shafts, bones, or teeth as nuclear DNA is not available for analysis or is too degraded [2]. Hair evidence without an intact root to provide adequate nuclear DNA contains plenty of mtDNA to be extracted and further analyzed [2]. While mtDNA cannot be used to identify a single source, mtDNA processing in cases with limited evidence or limited nuclear DNA can be used for exclusions. In the case of an inclusion, a haplotype frequency is determined using a population database to establish how rare or how common the particular haplotype is within a certain population [30]. Established databases of common sequence polymorphisms are available for the United States [30].

1.5 mtDNA Heteroplasmy

High mutation rates in mtDNA can cause some individuals to be deemed heteroplasmic, meaning they have more than one mtDNA type as seen in **Figure 1.4** [3]. The first investigation into mitochondrial heteroplasmy was conducted in 1983 with the results showing differences in mtDNA caused by nucleotide substitutions, insertions, or deletions [31]. Heteroplasmy has become increasingly more evident with improvements in technology, and individuals can be heteroplasmic in certain tissues while being homoplasmic (only one detectable mtDNA type) in others [3]. Such heteroplasmic tissues can vary between individuals, but the kidney, liver, and skeletal muscle tissues have been found to have higher rates of heteroplasmy [32]. Two types of heteroplasmy exist: point substitution and length heteroplasmy [3]. Point substitution heteroplasmy is a difference in a single base between two same-source samples within the HV1 or HV2 region [3].

Maternal inheritance of mtDNA heteroplasmy has been researched with up to 59% of heteroplasmies in offspring attributed to being transmitted by the mother as opposed to new mutations arising in the offspring [33, 34]. A famous example of heteroplasmy being transmitted from mother to offspring was in the identification of the remains of Tsar Nicholas Romanov [35]. Presence of point heteroplasmy in both the suspected remains of the Tsar as well as his brother Georgij Romanov confirmed the heteroplasmy in the brothers' maternal lineage as well as put to rest controversy around the Tsar's identification [35]. Studies have explored mtDNA genetic bottleneck as a mechanism for mtDNA heteroplasmy inheritance as heteroplasmy frequency has been found to change between generations [33, 36, 37].

The rate of heteroplasmy in different tissue types is an important factor that has been investigated by previous papers in regard to forensic interpretations of mtDNA. Differences in the mtDNA sequences of two samples from the same source can complicate the interpretation of the comparison as well as the report and testimony of the scientist [38]. Heteroplasmy in mtDNA has been studied using both traditional Sanger sequencing methods and next generation sequencing. Whole mitochondrial genome sequencing has also been completed using next generation sequencing to study heteroplasmy in the mtDNA genome [39]. Past studies have focused on the

entire mitochondrial genome or only the hypervariable regions of the control region [17, 40, 41]. Heteroplasmy in the three hypervariable regions of the mitochondrial genome has been studied with findings indicating that point heteroplasmy is more frequent in the HV1 and HV2 regions than the HV3 [40, 41]. The mtDNA control region from buccal swabs and blood samples was sequenced from 5,000 global samples in which point heteroplasmy was present in 6% of individuals across all samples [42].

Heteroplasmic occurrences have not been found to be significantly different between different populations or mtDNA haplogroups [39]. Heteroplasmic positions are more likely to occur in the control region of the mitochondrial genome than in other coding regions, but heteroplasmic positions can and do occur within mitochondrial genes [39]. Mitochondrial heteroplasmies in certain genes can be correlated to rare diseases and disorders [43, 44]. Several studies have demonstrated that mtDNA heteroplasmy present in hair is not always found in other sample types such as buccal swabs and blood [45–47]. Heteroplasmy in hair has previously been reported at rates up to 10.5% [45]. Gallimore et al. found 13% of hair samples taken either from the head or pubic region possessed heteroplasmy that was not present in buccal swabs or blood samples taken from the same participants [11]. Heteroplasmic positions have been also found to vary between hair shafts taken from the same individual, demonstrating that mtDNA heteroplasmy is not conserved between different hair shafts [48, 49].

Research relating heteroplasmy in hair samples has sometimes utilized nested-PCR in which two rounds of PCR are run prior to sequencing [50, 51]. In one such study by Grzybowski [51], high rates of point heteroplasmy was found in single hair roots. This study highlighted several issues with mtDNA processing that could lead to misleading results. Budowle et al. [52] critiqued the paper and cited the abnormally high levels of DNA template along with the exorbitant amount of PCR cycles used as reasons for the apparent heteroplasmy. Further research on heteroplasmy rates in hair samples is needed to aid in the interpretation of mtDNA comparisons, especially with the issues of contamination and processing methods playing such a large role in results and conclusions.

1.6 Current mtDNA Analysis

Like nuclear DNA testing, mtDNA testing must adhere to the FBI Quality Assurance Standards (QAS) outlined in the similarly named “Quality Assurance Standards for Forensic DNA Testing Laboratories” [53]. These standards dictate that proper documentation is kept of each step (QAS 7.1), with each lab having written analytical procedures for each method (QAS 9.1) [53]. mtDNA samples must follow similar typing procedures as nuclear DNA including extraction, purification, amplification, separation, and interpretation. A reagent blank must be subjected to the same extraction procedure at the time of sample extraction, and this reagent blank must be subjected to each of the following analytical procedures through separation [53]. While nuclear DNA samples must be quantified prior to amplification of the sample, mitochondrial DNA samples are not required to be quantified (QAS 9.4) [53]. SWGDAM still recommends that the FBI QAS are followed for mitochondrial analysis regardless of whether quantification is required for mtDNA [30]. Direct sequencing of mtDNA is typically performed utilizing Sanger sequencing with

resulting sequences being compared to a maternal lineage or the FBI mitochondrial DNA database [2, 28].

1.7 DNA Extraction

The first step in DNA analysis after confirming the presence of a biological fluid is to extract the DNA from the sample. Extracting the DNA is important as the protocol involves cells lysis and the separation of DNA from other cellular components such as proteins, lipids, and other elements that can interfere in upcoming steps [27]. Removing any potential inhibitors of the polymerase chain reaction (PCR) amplification is another goal of extraction that is essential to the successful generation of a profile from an evidentiary sample [27]. Common inhibitors that need to be removed are hemoglobin found in blood, indigo dye found in denim [27] and melanin found in in hair [2].

There are several methods of DNA extraction including organic extraction, Chelex extraction, solid-phase extraction, and FTA paper [2]. No matter the method, DNA extraction poses the largest risk of contamination of any step involved in DNA analysis [2]. Current methods of DNA extraction are frequently automated and involve solid-phase extraction through the use of either a silicon column or paramagnetic resin [2]. DNA interacts with the column or resin by binding to the media in certain conditions (e.g. salinity, pH) during washing steps [2]. Multiple washings are required of every extraction method (except Chelex extraction) to ensure all other cellular components and PCR inhibitors are removed before the extracted DNA sample moves on to the next step of quantitation [2]. Addition of an elution buffer allows the DNA to be eluted from the media with the eluate being what is used in future processing steps [27].

1.8 DNA Extraction from Hair

Extracting DNA from shed hair samples requires that the hair shaft be completely dissolved in lysis buffer to release all the DNA from the cells [8]. Proteinase K (ProK) and dithiothreitol (DTT) are required to aid in the lysis of hair shafts [8]. ProK acts to hydrolyze proteins [27] while DTT aids in disulfide reduction, eliminating the disulfide bonds in keratin [2, 54][54].

1.8.1 Promega™ Tissue and Hair Extraction Kit

The Promega™ Tissue and Hair Extraction Kit is designed to be utilized with the Promega™ DNA IQ™ System to extract DNA from a variety of tissues including hair, formalin-fixed tissue, and bone [8]. The DNA IQ™ System is a solid-phase extraction which utilizes silicon-coated paramagnetic resin as a binding agent [2, 55]. When in a solution with high-salt content, the paramagnetic resin binds to DNA in the sample [2]. As illustrated in **Figure 1.3,** the tube containing the sample is put on the MagneSphere® Technology Magnetic Separation Stand where the magnetic resin is pulled to the side of the tube [2, 55]. Multiple washes are conducted with the DNA being bound to the resin on the side of the tube to remove cellular components and potential

PCR inhibitors [2]. Once the DNA has been purified via the washes, an elution buffer of low-salt content is used to elute the DNA from the resin [2].

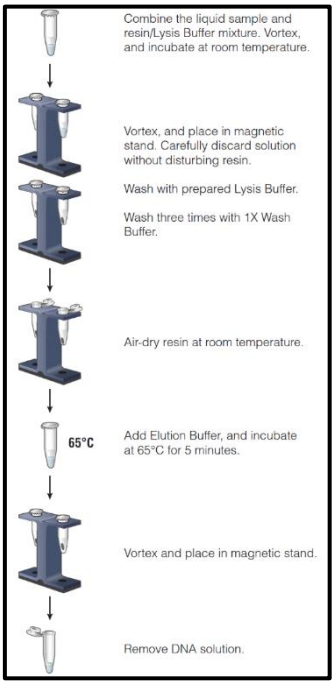

Figure 1.3: DNA IQ™ System [55]. *A schematic of the DNA IQ™ System with MagneSphere® Technology Magnetic Separation Stand. Silicon-coated paramagnetic resin is used to bind to DNA when in a high-salt solution. The MagneSphere® Technology Magnetic Separation Stand pulls the magnetic resin bound with DNA to the side of the tube while multiple washings take place to remove cellular components and purify the DNA.*

1.8.2 Qiagen QIAamp® DNA Investigator Kit

Qiagen manufactures several kits for DNA extraction. The QIAamp® DNA Investigator Kit is effective for the extraction of DNA from a variety of sources including hair roots and hair shafts [56]. While other extraction methods may require an ample sample size or volume, the QIAamp® DNA Investigator Kit is able to extract DNA from small volumes [56]. Regardless of the substrate, cells must first be lysed using an SDS-based detergent along with the addition of ProK for histone breakdown. DTT is also required for extraction for hair roots and shafts [56].

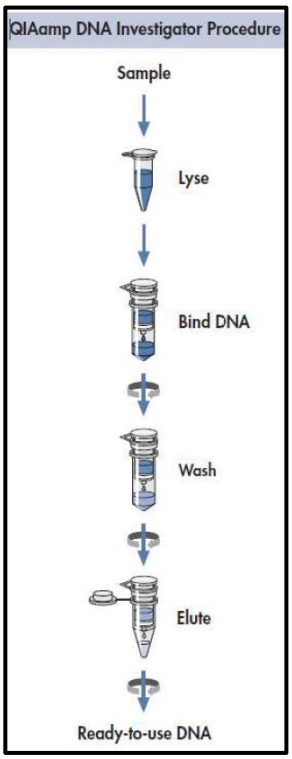

Figure 1.4: QIAamp® DNA Investigator Procedure [56]. *A schematic of the Qiagen QIAamp® DNA Investigator Kit workflow. After lysing cells, the solution is added to a silica-based column that selectively binds DNA in a choatropic salt solution. While bound, the DNA is washed using ethanol-based wash buffers before a low salt buffer is used to elute the DNA from the column.*

The lysate is transferred to a column with a silica-based membrane [56]. This membrane selectively binds DNA through the formation of salt bridges between the DNA and the membrane [56] In the presence of a chaotropic salt solution [56]. Ethanol-based wash buffers are added and spun through the column to wash away any contaminants and impurities from the DNA. A low-salt buffer is used to break the salt bridges and elute the DNA through the column after the washing steps have been performed, leaving purified DNA [56]. The entire workflow is illustrated in **Figure 1.4**.

1.8.3 Organic (Phenol-Chloroform) Extraction

Albeit more toxic than newer DNA extraction methods, the organic or phenol-chloroform extraction method is useful for extracting and purifying high molecular weight DNA [27]. SDS-based detergent, ProK, and DTT are once again used to lyse the cells and break down the histones and other proteins. Phenol-chloroform is added to the lysate as this mixture separates the DNA from the proteins and other unwanted cellular debris [27]. Upon centrifugation, the aqueous components separate from the organic components with the proteins and debris favoring the organic phase while the DNA is more soluble in the aqueous phase [27]. The upper aqueous phase containing the DNA can then be transferred to a new tube for the addition of more phenol-chloroform for additional removal of unwanted material [27]. DNA concentration can then be

performed with the use of centrifugal filters. This step not only reduces the overall volume of the extract, but also serves to remove salts from the extract [27].

1.9 Amplification and Sequencing of mtDNA

Amplification and sequencing of mtDNA requires two stages: amplification via PCR and cycle sequencing [27]. Similar to traditional nuclear DNA amplification, the primers selected are used to target a specific region of mtDNA. The PCR commonly targets HV1 and HV2 mtDNA regions. While commercial kits for nuclear DNA utilize fluorescently tagged primers, mtDNA amplification primers used are not tagged in such a way. This is followed by cycle sequencing where DNA polymerase is used to copy the template using one of two primers (either forward or reverse) [57]. This cycle sequencing step results in a linear and not exponential accumulation of product [57].

Success of the PCR amplification reactions can be determined through the use of agarose gel electrophoresis. The resulting bands can also be checked for size by comparison to the DNA ladder. While this method merely approximates the concentration of DNA in the PCR product, any issues with amplification can be caught before proceeding to the next step.

1.10 Sanger Sequencing

Sanger sequencing is the traditional method for sequencing with instrumentation requirements that are already available in current forensic DNA laboratories (**Figure 1.5**). After the targeted section of DNA is amplified in the first round of amplification, cycle sequencing is used to sequence that target. As with traditional Sanger sequencing, cycle sequencing uses dideoxy nucleotide triphosphates (ddNTPs) in addition to the typical dNTPs [2]. Each ddNTP is labelled with a different fluorescent dye to correspond to the base [2]. When the ddNTPs are incorporated during extension, no further bases can be added to the fragment being extended [2]. This is due to the structural differences of dNTPs and ddNTPs. dNTPs have a hydroxyl group at the 3´-position of the deoxyribose while there is only a hydrogen at this location on the ddNTPs [58]. Without the 3´-hydroxyl group, there is no possibility for the next nucleotide to be incorporated. Mixing the dNTPs and a lower concentration of ddNTPs results in the chain-terminating ddNTPs to create fragments of varying sizes depending on when the ddNTP is added.

A size based separation method such as capillary electrophoresis can then be utilized to effectively sort the differently sized fragments from smallest to largest [58]. After separation, a laser excitation source is used to excite the fluorophores attached to the ddNTPs [58]. Since the ddNTPs indicate the end of each chain, the detector reads the last nucleotide of each fragment in order of smallest to largest. A chromatogram is generated through the alignment of each final base. The placement of the peaks from left to right move from shortest fragments to longest fragments to reveal the order of the bases for the region targeted [2]. The resulting sequence can then be compared to other mtDNA sequences and databases to determine the likelihood of two individuals sharing the same mtDNA sequence [2].

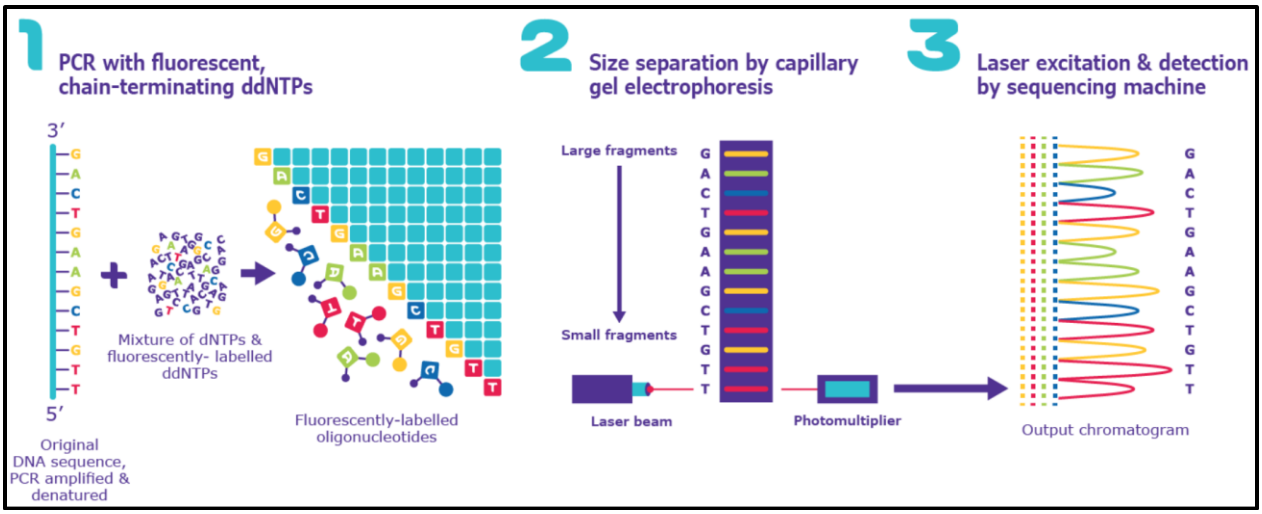

Figure 1.5: Sanger Sequencing Workflow [58]. *A visual representation of the workflow of Sanger sequencing is depicted from PCR to generation of a chromatogram.*

1.11 Purpose, Goals, and Objectives

This research aimed to accomplish two main goals: to successfully sequence the mtDNA control region from buccal swabs and to compare mtDNA between hair shafts and buccal swabs. The primary hypothesis was that mtDNA found in hair shafts may differ from buccal swabs and between hairs due to heteroplasmy being more prevalent in hair. Such differences can be recorded and used to better understand the rates of intra-source variability in mtDNA samples to aid in forensic interpretations. In addition to comparing mtDNA from the two sample types, three different extraction methods were used to compare their success extracting mtDNA from hair shafts.

2. Methods

2.1 Collection Process

Five willing participants were sampled after International Review Board (IRB) approval and informed consent was achieved. Participants were limited to female volunteers with long hair to ensure enough sample could be collected. To compare multiple samples from the same source, three sections of the scalp were differentiated. **Figure 2.1** depicts the three different hair sections, used for this research: Section 1 (front of head), 2 (crown of head), and 3 (nape of neck). 15-20 rootless strands of hair and 15-20 strands of hair containing the root were taken from each section. Each hair section was sampled in duplicate for quality control comparisons. Buccal swabs were also collected from each participant and were processed as reference samples.

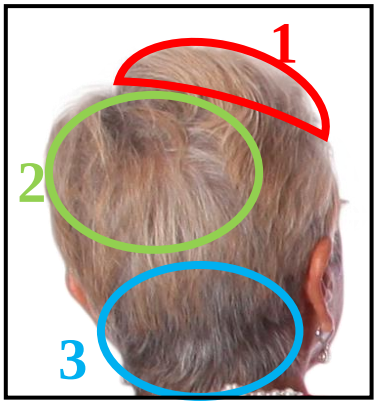

Figure 2.1: Hair sectioning. Sectioning of the scalp is depicted with three distinct areas being differentiated.

2.2 Buccal Swab DNA Extraction, Amplification, and Sequencing

Buccal swabs collected from each participant were extracted using the QIAamp® DNA Investigator Kit (Qiagen, Germantown, MD, USA) according to the manufacturer's guidelines for DNA isolation from buccal swabs. DNA extracts were quantitated through agarose gel electrophoresis.

The entire mtDNA control region (1,357 bp) was amplified via PCR with a total reaction volume of 25 µL using the F15851 and R639 primers (**Table 2.1**). Each reaction was comprised of 0.125 µL AmpliTaq Gold® DNA Polymerase (Thermo Fisher Scientific Inc., Carlsbad, CA, USA), 2.5 µL 10X PCR Buffer II, 3.0 µL 25 mM $MgCl_2$ (Thermo Fisher Scientific), 1 µL 10 mM dNTPs (Thermo Fisher Scientific), 0.5 µL of each 10 µM primer (F15851 and R639), and ~20 ng of DNA template. Amplification was performed on a GeneAmp™ PCR System 9700 thermal cycler (Thermo Fisher Scientific) under the following conditions: 95°C for 15 minutes; 30 cycles of 95°C for 30 seconds, 56°C for 1 minute, 72°C for 90 seconds; and a final extension at 72°C for 10 minutes. The PCR reactions were verified to be successful through gel electrophoresis. ExoSAP-IT™ (Thermo Fisher Scientific) was used to purify the PCR products according to the manufacturer's guidelines.

The cleaned PCR products were sequenced using eight internal primers (five forward primers and three reverse primers) to ensure both strands were sequenced (**Table 2.2**). Sequencing reactions were a total of 10 μL comprised of 2 μL BigDye™ Terminator 3.1 Ready Reaction Mix (Thermo Fisher Scientific), 1 μL 5X Sequencing Buffer (Thermo Fisher Scientific), 0.5 μL 10 μM either forward or reverse primer, and 3 μL of the purified PCR product. Cycle sequencing was performed on a GeneAmp™ PCR System 9700 (Thermo Fisher Scientific) under the following conditions: 96°C for 1 minute; 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 minutes. Sequenced products were cleaned using the BigDye™ XTerminator Purification Kit (Thermo Fisher Scientific) following the manufacturer's guidelines.

For separation of sequenced products, 10 μL of the cleaned product in HiDi formamide was loaded into a MicroAmp™ Optical 96-well reaction plate (Thermo Fisher Scientific) and separated via capillary electrophoresis using a 3500 Genetic Analyzer (Thermo Fisher Scientific) with the following specifications: POP-7™ Polymer (Thermo Fisher Scientific), 3500 Genetic Analyzer 8-capillary array, 36 cm (Thermo Fisher Scientific). The instrument protocol run module used was FastSeq36_POP7 (included in the Applied Biosystems® 3500 Series Data Collection Software) with no parameter changes made. The injection time was 8 seconds, and the injection voltage was 1.2 kV. Sequencing data was uploaded into the Sequencher® Sequence Analysis Software version 5.4.6 for Mac for sequence alignment and analysis.

Table 2.1: Amplification primers for the mtDNA control region as used by Chaitanya et al. [59].

Primer	Primer Sequence
F15851	5´- ATCTCCCTAATTGAAAACAAAATACTCAAA -3´
R639	5´- GGGTGATGTGAGCCCGTCTA -3´

Table 2.2: Sequencing primers for the mtDNA control region.

Primer	Primer Sequence
F16268	5´- CACTAGGATACCAACAAACC -3´
F15971	5´- TTAACTCCACCATTAGCACC -3´
F15851	5´- ATCTCCCTAATTGAAAACAAAATACTCAAA -3´
F314	5´-CCGCTTCTGGCCACAGCACT-3´
F15	5´-CACCCTATTAACCACTCACG-3´
R16	5´-TGATAGACCTGTGATCCATCGTGA-3´
R484	5´-TGAGATTAGTAGTATGGGAG-3´
R639	5´GGGTGATGTGAGCCCGTCTA-3´

2.3 Hair Sample Preparation and Extraction

For each hair sample, one inch of hair (either one inch including the hair root or one inch of the hair shaft) was cut into smaller sections 1-4 cm in length. Prepared hair samples from each scalp section of each participant were extracted using three extraction methods: Tissue and Hair Extraction Kit (Promega™, Madison, WI, USA), QIAamp® DNA Investigator Kit, and phenol-chloroform extraction.

Hair samples extracted using the Tissue and Hair Extraction Kit were processed following the manufacturer's guidelines for extracting mtDNA from hair follicles and hair shafts and eluted at 35 µL. Hair samples extracted using the QIAamp® DNA Investigator Kit were processed following the manufacturer's guidelines for extracting DNA from hair roots and hair shafts. These samples underwent a final room temperature incubation for 5 minutes as opposed to 1 minute and were eluted at 35 µL. Hair samples extracted using phenol-chloroform extraction were processed following the WVU Forensic Biology Laboratory protocol as shown in Appendix A. All samples were eluted with 35 mL of TE buffer.

2.4 Hair Extract Complete HV1 Amplification and Sequencing

The mtDNA HV1 region was amplified from hair extracts via PCR with a total reaction volume of 10 µL using the F15971 and R16410 primers (**Table 2.3**; product size 439 bp). Each reaction was comprised of 0.05 µL AmpliTaq Gold® DNA Polymerase (Thermo Fisher Scientific), 1 µL 10X PCR Buffer II, 1.2 µL 25 mM $MgCl_2$ (Thermo Fisher Scientific), 0.4 µL 10 mM dNTPs (Thermo Fisher Scientific), 0.2 µL of each 10 µM primer (F15971 and R16410), and 6.95 µL of DNA extract. Amplification was performed on a GeneAmp™ PCR System 9700 thermal cycler (Thermo Fisher Scientific) under the following conditions: 95°C for 10 minutes; 40 cycles of 95°C for 30 seconds, 56°C for 1 minute, 72°C for 90 seconds; and a final extension at 72°C for 10 minutes. The PCR reactions were verified to be successful through gel electrophoresis. The successfully amplified products were purified using ExoSAP-IT™ (Thermo Fisher Scientific) according to the manufacturer's guidelines.

The cleaned PCR products were sequenced in separate reactions using the same forward and reverse primers used for amplification (**Table 2.3).** Sequencing reactions were a total of 10 µL comprised of 2 µL BigDye™ Terminator 3.1 Ready Reaction Mix (Thermo Fisher Scientific), 1 µL 5X Sequencing Buffer (Thermo Fisher Scientific), 0.5 µL 10 µM primer, and 3 µL of the purified PCR product. Cycle sequencing was performed on a GeneAmp™ PCR System 9700 (Thermo Fisher Scientific) under the following conditions: 96°C for 1 minute; 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 minutes. Sequenced products were cleaned using the BigDye™ XTerminator Purification Kit (Thermo Fisher Scientific) following the manufacturer's guidelines. For size separation of sequenced products, 10 µL of the cleaned product in Hi Di formamide was loaded into a MicroAmp™ Optical 96-well reaction plate (Thermo Fisher Scientific) and separated via capillary electrophoresis using a 3500 Genetic Analyzer (Thermo Fisher Scientific) with the following specifications: POP-7™ Polymer (Thermo Fisher Scientific), 3500 Genetic Analyzer 8-capillary array, 36 cm (Thermo Fisher Scientific). The instrument

protocol run module used was FastSeq36_POP7 (included in the Applied Biosystems® 3500 Series Data Collection Software) with no parameter changes made. The injection time was 8 seconds, and the injection voltage was 1.2 kV. Sequencing data was uploaded into the Sequencher® Sequence Analysis Software version 5.4.6 (Gene Codes Corporation, Ann Arbor, MI, USA) for Mac for sequence alignment and analysis.

Table 2.3: Primers for amplification and cycle sequencing of the mtDNA HV1 region.

Primer	Primer Sequence
F15971	5´- TTAACTCCACCATTAGCACC -3´
R16410	5´- GAGGATGGTGGTCAAGGGA -3´

2.5 Hair Extract Partial HV1 Amplification and Sequencing

The mtDNA HV1 region was amplified from hair extracts via PCR with a total reaction volume of 10 μL using in two parts. The first half of HV1 (280 bp) was amplified using the F15971 and R16251 primers (**Table 4**) and is henceforth referred to as HV1 5´. The second half of HV1 (266 bp) was amplified in separate reactions from HV1 5´ using the F16144 and R16410 primers (**Table 2.4**) and is henceforth referred to as HV1 [3´]. Each reaction was comprised of the same reagents and quantities as described in Section 2.4 with the appropriate primer pairs added. Amplification and cycle sequencing were performed using the same methods as described in Section 2.4. Sequencing data was uploaded into the Sequencher® Sequence Analysis Software version 5.4.6 (Gene Codes) for Mac for sequence alignment and analysis.

Table 2.4: Primers used for both amplification and cycle sequencing of the mtDNA HV1 [5´] and HV1 [3´] regions.

Region	Primer	Primer Sequence
HV1 [5´]	F15971	5´-TTAACTCCACCATTAGCACC-3´
HV1 [5´]	R16251	5´-GGAGTTGCAGTTGATGT-3´
HV1 [3´]	F16144	5´-TGACCACCTGTAGTACATAA-3´
HV1 [3´]	R16410	5´-GAGGATGGTGGTCAAGGGA-3´

2.6 Data Analysis

All sequences were uploaded to the Sequencher® Sequence Analysis Software version 5.4.6 (GeneCodes) for Mac. Forward and reverse reactions for each sample were aligned using the software. Coverage of both strands for the entire control region was confirmed for each participant's buccal swab samples. Hair shaft samples for HV1 [5´] and HV1 [3´] were aligned and only samples with coverage of both strands were considered successful.

The revised Cambridge Reference Sequence (rCRS) [29] was uploaded and aligned to each sample to ensure the correct mtDNA base positions were used to describe the sample sequences. Each instance of divergence from the rCRS was investigated to determine if the difference was due to an error in base calling as opposed to a true polymorphism from the rCRS. All polymorphisms were recorded for each participant. The haplogroup of each participant was determined by uploading the buccal swab sequences to Foswiki Mitomaster [60]. All ambiguous base calls for each uploaded sequence were investigated in Sequencher® to determine instances of two different base calls at the same location. Instances of point heteroplasmy were verified by noting the presence of two bases at that position on all forward and reverse strands for those samples. Any samples with point heteroplasmy were aligned with and compared to the reference buccal swab and any other successfully sequenced hair samples from that participant. Length heteroplasmy was verified by confirming the presence of out-of-phase chromatograms in the appropriate sections on both the forward and reverse strands for the sample.

3. Results and Discussion

3.1 Polymorphism and Haplogroup Determination

Polymorphisms identified in the control region for each participant are shown in **Table 3.1**. The presence of each polymorphism was confirmed in both strands as well as in sequenced hair shaft samples. As the presence of certain polymorphisms are the basis for the identification of mtDNA haplogroups and haplotypes, the sequence of each participants' control region was uploaded to Mitomaster [60].

Table 3.1: Polymorphisms from rCRS in each participant.

	Position																								
Participant	16223	16240	16278	16282	16292	16294	16309	16311	16325	16356	16390	16519	73	146	152	189	195	200	263	309*	315*	497	499	524*	525*
1	T	-	T	T	-	T	C	-	-	-	A	C	G	C	C	-	C	-	G	C	C	-	-	-	-
2	T	G	-	-	T	-	-	C	-	-	-	C	G	-	C	G	-	G	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	C	-	-	-	C	G	-	-	-	-	-	G	-	C	T	-	A	C
4	-	-	-	-	-	-	-	-	C	-	-	C	-	-	-	-	-	-	G	C	C	-	-	-	-
5	-	-	-	-	-	-	-	-	-	C	-	C	G	-	-	-	C	-	G	-	-	-	A	A	C

** Indicates the base shown was an insertion after the listed position.*

Table 3.2 shows the assigned haplogroups. Two participants belong to haplogroup L which is the oldest haplogroup showing humankind's African roots [21]. Haplogroup L3 in particular led to haplogroups M and N which are known to represent groups that migrated out of Africa with haplogroup N leading to subclade R [20]. Both haplogroup H and U are lineages of haplogroup R and are common in western Eurasia [61]. Haplogroup K in turn is a lineage of U (specifically, haplogroup U8) [20].

Table 3.2: Participant Haplogroups. Haplogroups assigned by Mitomaster. The consensus reference sequence for each participant was uploaded. Haplotypes included in brackets.

Participant	Haplogroup
1	**L** [L2a1a2]
2	**L** [L3f1b + 16292]
3	**K** [K1a]
4	**H** [H2a2a]
5	**U** [U4]

3.2 Hair Shaft PCR Amplification Success

3.2.1 Control Region

Sequencing of the mtDNA control region was largely successful using buccal swab extracts. All reference sequences for the entire control region were successfully sequenced using just one forward and one reverse primer for amplification followed by five forward and three reverse primers during cycle sequencing. This led to the ability to align all sequencing reactions in Sequencher® to verify each base in multiple overlapping strands.

Attempting to sequence the control region in the same manner using hair shafts was not met with the same success. No successful amplified products were visualized using gel electrophoresis for any hair shaft sample (n = 45). Even when increasing the number of amplification PCR cycles to 40 there were no successful products. The size of the amplicon (1,357 bp) was determined to simply be too large to be amplified as one fragment for hair samples. Previous research on sequencing mtDNA from hair targeted much smaller products. Allen et al. amplified products 567 bp and 520 bp [50]. Another study targeted even smaller overlapping fragments between 176-409 bp in buccal swabs, bone, and hair [62]. Degraded DNA sometimes requires the use of mini primer sets. Gabriel et al. used such mini primers sets to target fragments 126-170 bp in length when studying mtDNA sequencing from dried skeletal remains [63].

3.2.2 HV1

After the failure to amplify the entire control region from hair shafts, an adjustment was made to target only HV1 (439 bp). Only four hair samples across all five participants were successfully amplified (n = 45) showing a 9% success rate. If the 439 bp fragment failed to amplify, two smaller overlapping fragments, HV1 [5´] (280 bp) and HV1 [3´] (266 bp) were amplified to improve the success of amplification. This change resulted in ten successfully amplified products for HV1 [5´] and twenty successfully amplified products for HV1 [3´] with success rates of 11% and 22% for each region, respectively (n = 90 for each HV1 region). As not all amplified products yielded sequencing results for both the light and heavy strand, overall success of samples from extraction to sequencing were also calculated. Overall sequencing success rates for HV1 [5´] and HV1 [3´] were 4% and 15%, respectively. Success of each HV1 region is shown in **Table 3.3**.

Table 3.3: Comparison of HV1 region success rates from hair shafts (n = 90 for each HV1 [5´] and HV1 [3´]).

	HV1 Region	
	5´ [280 bp]	*3´ [266 bp]*
# Successfully amplified	10	20
Success rate (ext-amp)	11%	22%
# Successfully sequenced	4	14
Success rate (ext-seq)	4%	15%

Additionally, the success rates of each extraction method were calculated to compare the efficacy of each kit for the extraction of mtDNA from hair shafts. The number of successfully amplified samples was recorded for each extraction method and each HV1 region (5´ or 3´). Success rates for each extraction method are shown in **Table 3.4** with the rates for success from extraction to amplification as well as from extraction to sequencing included.

Table 3.4: Comparison of extraction method success rates from hair shafts (n= 30 for each extraction method). The bracketed numbers show the number of samples with HV1 [5´] targeted followed by the number of samples with HV1 [3´] targeted. Only samples with coverage of both light and heavy strands were considered successful.

	Extraction Method		
	*Promega™**	*Qiagen**	*Organic*
# Successfully amplified	12 [5,7]	4 [0,4]	15 [5,10]
Success rate (ext. to amp.)	40%	13%	50%
# Successfully sequenced	6 [3,3]	4 [0,4]	8 [1,7]
Success rate (ext. to seq.)	20%	13%	27%

* Shortened names for each extraction method were used above. The Promega™ extraction kit used was the Tissue and Hair Extraction Kit following the manufacturer's protocol for mtDNA from hair. The Qiagen extraction kit used was the QIAamp® DNA Investigator Kit following the manufacturer's protocol for extraction from hair shafts.

Out of the eighteen samples successfully sequenced there were seven from Participant 1, six from Participant 2, one from Participant 3, three from Participant 4, and one from Participant 5. This uneven distribution of success indicates that the ability to extract and amplify mtDNA from hair shafts may be due to differences between participants such as hair type and mtDNA copy number. Desmyter et al. [64] demonstrated that mtDNA sequencing success can vary based on the location of the hair shaft cutting to the hair root with 96% of hair shafts taken 1 cm from the root resulting in the control region being successfully sequenced. The same study also reported 51% control region sequencing success in a longitudinal study in which twenty 2 cm fragments from each hair shaft were taken moving away from the root [64]. Other studies have also shown that mtDNA copy number generally decreases and mtDNA degradation increases as cuttings of hair farther from the root are used [65–67]. As the proximity to hair root was not investigated in this research, the location along the hair shaft was not recorded during sampling or extraction preparation and could have played a role in the sequencing success of the hair shafts. mtDNA copy number in hair can also vary between individuals as well as between different hair shafts taken from the same individual [65, 66]. mtDNA quantities from extracts were not determined, so copy number may affect the amplification success between samples and individuals.

Regarding hair type, studies have reported conflicted results on whether hair color or diameter affects amplification or sequencing success [68, 69]. Melton et al. [68] found the likelihood of obtaining a mtDNA profile from hair was higher for hair samples that were darker in color and

larger in diameter. A later study by Roberts and Calloway [69] found that amplification success from hair shafts did not have any correlation to hair color, hair diameter, or hair treatment (i.e. bleaching, coloring). Melanin has been previously reported to be a PCR inhibitor as it binds to DNA [70]. Studies have shown that inhibition due to melanin affects PCR efficiency of larger amplicons (>300 bp) than smaller amplicons (<230 bp) [70, 71]. Even so, Melton et al. [68] found melanized hair to yield abundant mtDNA. The samples used in the current study were anonymized to meet IRB requirements so no conclusions about melanin and its effects on PCR efficiency were reported.

AmpliTaq Gold® DNA polymerase used in the current study has recently been compared to eleven other DNA polymerases on the market [72]. When the entire control region was amplified from hair samples, AmpliTaq Gold® was outperformed by five other polymerases [72]. The same study also found that amplifying head hair resulted in lower mtDNA yields than beard/pubic hair [72]. As head hair was amplified using AmpliTaq Gold® in the current study, this combination may have affected amplification success. Additionally, chemically treated head hair (i.e. bleaching, coloring, chemical straightening) resulted in lower mtDNA yields compared to untreated head hair contrary to the Roberts and Calloway findings when determining amplification success [69, 72]. Amplification success in the current study may have been affected by hair treatment as participants were not asked about chemical treatments prior to sampling.

A validated Sanger sequencing method for HV1 and HV2 utilized one of the same extraction methods (Qiagen QIAamp® DNA Investigator Kit) as the current study and sequenced hair shafts [62]. All five hair shafts used in this validation study were all successfully sequenced, and the study targeted smaller overlapping fragments (176-490 bp) during amplification [62]. The major differences between Mita et al. [62] and the current study was the former's use of Terg-A-Zyme for hair shaft cleaning prior to extraction and the use of different primers for mtDNA amplification. Additionally, Mita et al. [62] utilized AmpliTaq Gold® Fast PCR Master Mix. The differences in the primers and DNA polymerase used between the current study and Mita et al. [62] may explain the difference in success.

3.3 Heteroplasmy

3.3.1 Point Heteroplasmy

Point heteroplasmy was found in the buccal swab (n = 5) and one hair sample (n = 18) from the same participant (Participant 3). Chromatograms from Participant 3's hair buccal swab and hair shaft can be found in Appendix B. **Table 3.5** shows the comparison of point heteroplasmy positions to the reference sequence of the participant. The buccal swab showed heteroplasmy at only one position (16093), but this position was not within the HV region sequenced in the hair shaft and therefore could not be verified in the hair shafts from the same participant. Position 16093 has been previously reported to have a high mutation rate resulting in heteroplasmy. Heteroplasmy at position 16093 has been reported in other studies involving blood samples and

hair samples [42, 68] Position 16093 has been found to not only be an evolutionary mutational hotspot, but has also demonstrated to have an increased rate of mutation in individuals exposed to radiation [73].

In the current study, point heteroplasmy was found in one out of five participants (Participant 3). An estimation of point heteroplasmy rate could not be determined due to the small number of successful samples. Heteroplasmy in the hair shaft from Participant 3 was seen at two different positions (16258 and 16288) in the HV1 [3´] region. No heteroplasmy was seen in the participant's buccal swab at these positions. Past studies have reported heteroplasmy in hair that is not seen in blood or saliva samples from the same individual [47, 74]. Heteroplasmy in hair has been reported at rates between 4.4-7% in HV1 and HV2 [46, 47]. While having more than one position showing heteroplasmy is uncommon, several studies have reported individuals showing heteroplasmy at two or more positions. Out of 35 individuals showing heteroplasmy, Tully et al. [75] found two individuals to each be heteroplasmic at two positions when sequencing HV1 in blood samples. Irwin et al. [42] found 0.14% of blood samples and buccal swabs to show heteroplasmy at two or three positions. A more recent study sequencing the whole mtDNA genome from hair reported out of thirteen participants showing heteroplasmy in hair, seven of them showed heteroplasmy at more than one position [74].

Table 3.5: Heteroplasmy in samples. Hair shaft sample was collected from Section 2 of the scalp and extracted using an organic extraction.

Participant	Sample type	Sequenced region	Heteroplasmy position	Bases called	Participant reference
3	Hair shaft	HV1 [3´]	16258	A/G	A
			16288	C/T	T
3 Ref.*	Buccal swab	Control region	16093	T/C	T/C

* Reference buccal swab taken from participant and extracted using the Qiagen QIAamp® DNA Investigator Kit following the manufacturer's Surface and Buccal Swab Protocol.

While each participant had three different scalp regions sampled for hair shafts, no other samples were successfully sequenced for Participant 3, so variability in heteroplasmy between scalp regions could not be assessed. In terms of interpretation when comparing Participant 3's buccal swab to the hair sample, one base present in the hair sample was concordant with the buccal swab base call at each position. Following the SWGDAM guidelines for mtDNA analysis interpretation [30], the samples would not be excluded as coming from the same source. As no other hair samples from Participant 3 were successfully sequenced, it is unknown whether any hair shaft would show a single base at the present heteroplasmic positions. A hair shaft showing the non-concordant base when compared to the buccal swab at both positions would indeed result in a false exclusion based on the SWGDAM guidelines [30].

3.3.2 Potential Length Heteroplasmy

No length heteroplasmy was seen in any hair sample sequenced. Two participants (Participants 4 and 5) showed potential length heteroplasmy in their buccal swabs. For both participants, this heteroplasmy was seen at the HV2 poly-cytosine (poly-C) stretch at positions 303-315 of the light strand. Out-of-phase chromatograms were seen after this position for both the light and heavy strands (**Figure 3.2**). Length heteroplasmy at this particular stretch is common and can be identified by the presence of out-of-phase chromatograms after position 309 on the light strand and before position 309 on the heavy strand [76, 77]. Participant 4 shows a combination of 8 cytosine repeats and 9 cytosine repeats beginning at position 303 on the light strand. Participant 5 shows a combination of 7 cytosine repeats and 9 cytosine repeats beginning at the same position.

Length heteroplasmy has been reported in all three HV regions in previous studies. Rasmussen et al. [41] saw length heteroplasmy in 19% of individuals sampled in either HV1, HV2, or both regions. Paneto et al. [40] focused on heteroplasmy in HV3 and found one participant showed length heteroplasmy in the CA repeat region in both blood and hair samples. Three participants in the Paneto et al. [40] study also showed length heteroplasmy at position 573 (poly-C stretch).

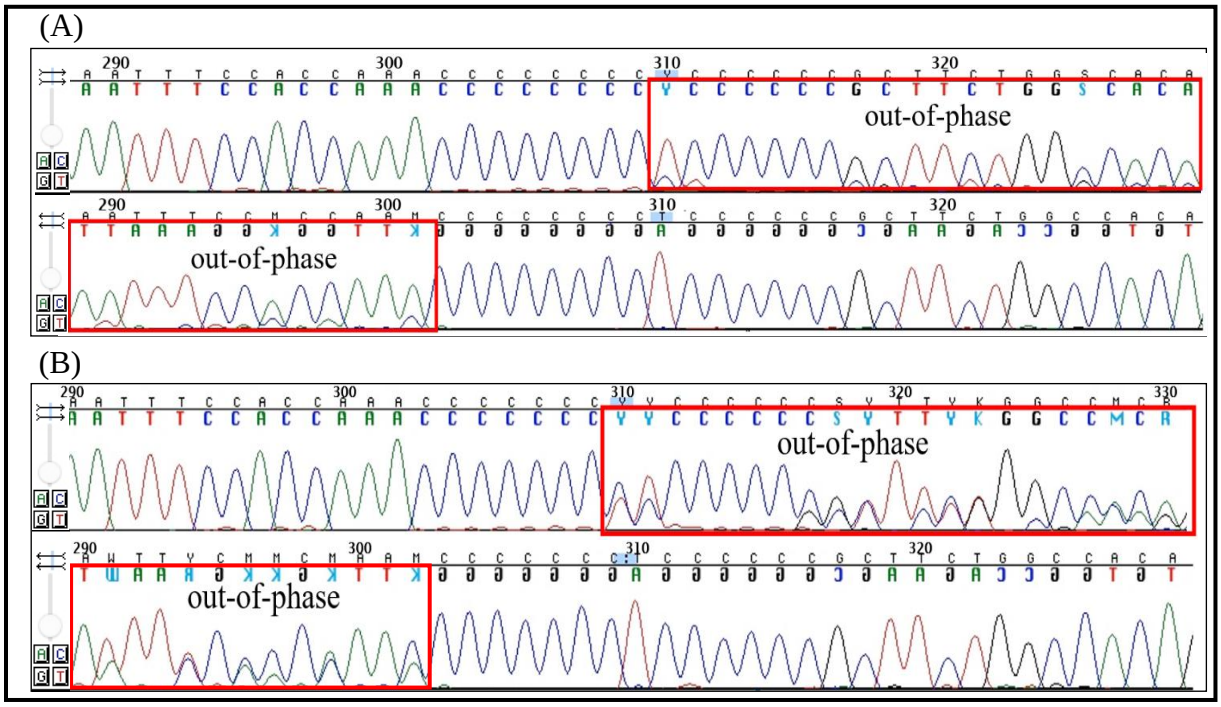

Figure 3.2: Chromatograms for Participants 4 (A) and 5 (B) showing potential length heteroplasmy. Top chromatograms for A and B show the light strand while the bottom chromatograms show the heavy strand.

No length heteroplasmy was seen in HV1 in any buccal swab or hair sample in this study, but the low success rates seen in the hair samples (particularly in HV1[5´]) prevented the entirety of HV1

from being sequenced in most hair samples. A challenge with interpreting length heteroplasmy is that polymerase slippage at homopolymeric stretches can also result in what seems like heteroplasmy. Length heteroplasmy is consequently disregarded in forensic interpretation. Berger et al. [78] remarked upon the difficulty in interpreting Sanger sequencing data in which length heteroplasmy is present. One study found that 52% of 5,015 individuals showed length heteroplasmy somewhere within the control region [42]. Presence of length heteroplasmy results in difficult-to-interpret chromatograms that may be mistaken for mixed samples and complicates sequence interpretation.

3.4 Importance of Sequence Comparison

mtDNA sequencing requires a clean laboratory space to reduce the risk of contamination [79]. Due to the high copy number of the mtDNA genome per cell versus that of nuclear DNA, even contamination involving few cells can result in a mixed sample when targeting mtDNA. Cleaning and washing of hair samples prior to extraction removes exogenous DNA from the hair, and different washing methods have been studied to limit the amount of contamination from such exogenous DNA [80, 81]. Studying heteroplasmy relies on the assumption of uncontaminated samples so the presence of more than one base at a single position can be attributed to the individual's actual sequence and not a mixture of one or more contributors. In the event of point or length heteroplasmy in this research, comparison of sequences containing the heteroplasmy to other participants' samples proved helpful in identifying contamination as the source of the heteroplasmic positions.

3.4.1 Assessing Point Heteroplasmy

One hair shaft sample (HV1 [3´] sequenced) in this study showed heteroplasmy at nine different positions that were not heteroplasmic in any of three other hair shaft samples processed from the same participant. While an objective of this study was to determine if hair shafts from different scalp sections could show differences in heteroplasmy, for one sample to have nine different heteroplasmic positions in a target barely longer than 260 bp would have been unprecedented. Grzybowski [51] previously reported high rates of point heteroplasmy in hair when sequencing HV1, and one individual in this study had heteroplasmy at six different positions. A critique of Grzybowski's [51] paper explained the heteroplasmy seen at two or three positions are possible but occur at very low rates [52]. As six heteroplasmic positions is considered higher than anything previously reported, a thorough investigation into the hair shaft sample with nine such positions was conducted.

The hair shaft sample was aligned with not only the reference buccal swab sequence for its participant (Participant 1), but for every other participant as well. This was done to determine (1) which base at each heteroplasmic position could be attributed to the expected participant's buccal swab and (2) which of the other participants could have contributed the second base at each

position. Only one participant (Participant 2) who did not contribute the hair shaft in question could have contributed the second base at each heteroplasmic position based on their buccal swab sequence. To account for any differences that may be in the sequences of the buccal swab and hair shaft, an additional comparison was made between the heteroplasmic hair shaft from Participant 1 and a hair shaft processed from Participant 2. The results of this comparison are shown in **Table 3.6**. Comparison of the hair shaft from Participant 1 with Participant 2 hair samples showed no locations where the participants' sequences differed apart from the heteroplasmic positions seen in Participant 1 hair shaft. This indicates that the "heteroplasmy" is merely due to contamination by a sample from Participant 2.

Table 3.6: Investigation of Participant 1 hair shaft heteroplasmy. Participant 2 could have contributed the second base at each position with one base called.

*Hair shaft collected from scalp section 2 and was extracted via the Qiagen QIAamp® DNA Investigator kit following the manufacturer's Hair Shaft protocol. Only HV1 [3´] was successfully sequenced.

	Participant 1		**Participant 2**	
Heteroplasmy position	*Reference*	*Hair shaft**	*Reference*	*Hair shafts†*
16209	T	C/T	C	C
16240	A	A/G	G	G
16278	T	C/T	C	C
16286	T	C/T	C	C
16292	C	C/T	T	T
16294	T	C/T	C	C
16309	G	A/G	A	A
16311	T	C/T	C	C
16390	A	A/G	G	G

†All hair shafts from Participant 2 with HV1 [3´] successfully sequenced were aligned for comparison and all shared the same base at each position. Five hair shafts were included: three extracted via phenol-chloroform extraction (one each from the front of head, crown of head, and nape of neck) and two extracted via the Qiagen QIAamp® DNA Investigator kit following the manufacturer's Hair Shaft protocol (one from the front of head and one from the nape of neck).

As the methods of this study did not include cleaning of the hair shafts prior to extraction, this contamination may have been due to the presence of exogenous DNA from Participant 2 on the hair shaft that was subsequently extracted along with the mtDNA found within the hair. The peak height ratio of major to minor heteroplasmic peaks ranged from 44% (positions 16278 and 16294) to 72% (position 16240). Due to the amount of exogenous DNA that would need to have been present to result in such ratios, a better explanation for the contamination would be analyst error. All hair samples with HV1 [3´] amplified were processed together. Therefore, a contamination

event due to analyst error is a possible explanation that could have occurred at any step from extraction through sequencing.

A similar investigation was conducted to assess the possibility of contamination playing a role in the heteroplasmy found in the Participant 3 hair shaft where two heteroplasmic positions were observed. No other participant could have contributed the second base at either heteroplasmic position when the reference buccal swab and hair shaft samples were compared. Additionally, the buccal swab sequence from the analyst was compared and was found that this sequence also could not account for the second base found at either heteroplasmic position.

3.4.2 Assessing Potential Length Heteroplasmy

Potential length heteroplasmy was seen in Participants 4 and 5, with a different number of repeated cytosines beginning at position 303 in the HV2 region. Length heteroplasmy at this position has been previously reported in two homopolymeric regions (C-stretches) within the mtDNA control region, one within the HV2 region between positions 303-315 and another within the HV1 region between positions 16184-19193 [76, 81]. If the length heteroplasmy came from contamination by another sample, we would expect to see point heteroplasmy at positions that differ between each participant and the second contributor. Comparison of Participants 4 and 5 reference buccal swabs to each of the other participants' reference buccal swabs showed multiple positions in which the sequences differed from the other participants, yet no point heteroplasmy was seen for Participants 4 or 5. All sequence differences between participants can be found in Appendix A. This supports that the length heteroplasmy in both participants is indeed a result of a combination of mtDNA containing a different number of repeated cytosines.

While contamination was investigated as a possible source of the out-of-phase chromatograms, *Taq* slippage during PCR amplification was not ruled out in the current study. *Taq* polymerase slippage during PCR can cause sequence errors such as a combination of different poly-C repeats in one sample [82–84]. Nakai et al. [84] found heteroplasmy between nucleotide positions 303-315 that were confirmed through the use of varying primers along with proof-reading polymerases.

4. Conclusions

Success in sequencing the control region of mtDNA from buccal swabs and HV1 from hair shafts showed that differences between sample types due to heteroplasmy is possible. While the small sample size prevents heteroplasmy rates from being estimated, appearance of length and point heteroplasmy were able to be reported. Two out of five unrelated participants (40%) showed potential length heteroplasmy in the 303-315 poly-C region of HV2 when mtDNA from buccal swabs was sequenced. One out of five unrelated participants (20%) showed point heteroplasmy at position 16093 (HV1) when mtDNA from their buccal swab was sequenced. Only one hair sample showed point heteroplasmy within HV1 (nucleotide positions 16258 and 16288). Neither position was heteroplasmic in the buccal swab from the same participant.

Past studies have shown higher success in sequencing HV1 from hair shafts than was seen in this research. There seemed to be a difference in the success rates for the HV1 sections sequenced with only 11% of HV1 [5´] samples being successfully amplified while the success rate for amplification of HV1 [3´] was 22% (n = 90 for each section of HV1). Such a disparity between the success rates may be due to DNA quality from the hair shafts as targeting a smaller amplicon size showed higher success. Future studies recording distance of the hair cutting from the root can help improve PCR amplification from hair shafts as hair cutting location has been correlated to mtDNA copy number and degradation. Furthermore, studying the effects of hair treatments such as bleaching and dyeing on the efficacy of mtDNA amplification would be helpful as such factors were not reported for the participants of the current study. Applying such methods to tissues other than buccal swabs would require additional protocol development and optimization to increase the success rate of amplification.

While three different methods of extraction were used, it was possible to successfully sequence mtDNA using each kit. Although all extraction methods used in this study involved purification of the extracted mtDNA, washing the hair prior to extraction may decrease the chances of sample contamination as one hair sample sequenced showed contamination from another participants' mtDNA. Optimizing a method in which more samples can be sequenced for additional comparison would increase the weight of future data related to heteroplasmy rates in hair compared to other sample types. Sampling from an increased number of scalp sections would also allow for additional comparison to determine the rate in which heteroplasmy may be expressed in only some hair shafts collected from a single individual.

Even with the low success rates reported in this study, the importance of research involving mtDNA heteroplasmy and its potential effect in forensic sample interpretation is evident. Comparing the hair sample with point heteroplasmy to the same participant's buccal swab revealed the buccal swab sequence did not share heteroplasmy at either position. While the hair sample would not be excluded as coming from the same source as the buccal swab, no other hair samples from the participant were successful. Given the inability of this study to compare multiple hair samples from the participant, a question is raised as to whether some hair shafts may only show the presence of the non-concordant base. If this were the case, then a sample with two non-concordant positions to the reference buccal swab would justify a false exclusion. Without the success of additional hair shaft samples taken from other scalp sections of the participant, this

question could not be answered in this study. Nevertheless, continued research in identification and comparison of heteroplasmy in hair shafts, buccal swabs, and other sample types is paramount to the strengthening of mtDNA analysis and interpretation in forensic science.

5. Appendix A: Protocols

Organic Extraction Protocol:

1. Place the sample cutting in a sterile 1.5 mL microcentrifuge tube.
2. Add 500 µL of stain extraction buffer, 15 µL of ProK, and 50 µL DTT.
3. Pulse vortex for 10 seconds and briefly centrifuge (8,000 rpm for 1 minute).
4. Incubate the sample at 56°C for 2-24 hours or until the hair is completely broken down.
5. Centrifuge at 8,000 rpm for 1 minutes to force condensation to the bottom of the tube.
6. In a fume hood, add 500 µL of Phenol Chloroform : Isoamyl Alcohol to the stain extract. Pulse vortex to attain a milky emulsion.
7. Spin the tube at 14,000 rpm for 3 minutes. There should be two clearly separated layers.
8. Remove the bottom organic phase and the interface from the tube and discard into a dedicated waste container.
9. Repeat steps 6-8 until middle layer is clear (at least 2 times).
10. Place a Microcon® 30kDA Centrifugal Filter Unit (Millipore Sigma, St. Louis, MO, USA) into a microcentrifuge tube.
11. Add 100 µL of TE buffer to the Microcon®.
12. Transfer the top aqueous phase from the tube in step 9 to the Microcon®. Avoid pipetting any of the bottom organic phase or interface. Discard the tube containing the organic phase and interface into a dedicated waster container.
13. Cap the Microcon® and centrifuge at 14,000 rcf for 10 minutes. Repeat until all of the solution flows through the Microcon®.
14. Carefully remove the Microcon® from the microcentrifuge tube and discard the fluid from the tube. Return the Microcon® to the tube.
15. Add 200 µL of TE buffer to the Microcon®. Cap and centrifuge at 14,000 rcf for 10 minutes. Repeat until all the solution flows through the Microcon®.
16. Add TE buffer to the Microcon® to bring the total volume between 20-200 µL. Carefully invert the Microcon® into a fresh, sterile microcentrifuge tube.
17. Centrifuge at 1,000 rcf for 3 minutes.
18. Store the sample at 5°C or frozen. When ready to use, vortex and centrifuge for 5 seconds.

DNA Amplification Protocol

Store all reagents on ice while preparing reactions

1. Calculate the volume of DNA extract of each sample needed to add 20 ng of total DNA
   NOTE: Maximum volume of added DNA is 17.375 µL, so extract may need to be diluted or concentrated
2. Determine number of reactions needed by adding number of samples, number of extraction blanks, one positive control, and one negative control
3. Add 4 to the calculated number of reactions to ensure enough master mix is made. This value is your N for the next step
4. Create the master mix by vortexing each of the following components and mixing the appropriate amount of each in a tube:
   a. N x 0.125 µL AmpliTaq Gold® DNA Polymerase
   b. N x 2.5 µL 10X PCR Buffer II
   c. N x 3.0 µL 25 mM $MgCl_2$
   d. N x 1.0 µL 10 mM dNTPs
   e. N x 0.5 µL 10 µM forward primer*
   f. N x 0.5 µL 10 µM reverse primer*
5. Pulse vortex the master mix for ten seconds and briefly centrifuge to remove any bubbles
6. Dispense 7.625 µL of master mix into each PCR tube
7. Vortex and briefly centrifuge DNA extracts. Add the volume of DNA extract calculated in Step 1 to target 20 ng of total DNA.
8. Add the appropriate volume of 0.1 TE buffer to bring each reaction volume to 25 µL
9. Pulse vortex each reaction for 10 seconds and centrifuge for 20-30 seconds.
10. Ensure there are no bubbles and perform a pre-PCR volume check
11. Load the reactions on a thermal cycler and run with the following parameters:

Hold	Cycling (30 cycles)			Final Extension	Hold
95°C	95°	56°C	72°C	72°C	12°C
15 min.	30 sec.	1 min.	90 sec.	10 min.	∞

12. Store PCR products at –20°C until ready for PCR product cleanup

*Forward and reverse primers will vary based on the region of DNA being targeted. Only add one forward and one reverse primer to the master mix

PCR Product Cleanup (if sequencing)

Remove ExoSAP-IT™ reagent from –20°C freezer and keep on ice throughout this procedure

1. Add 2 µL ExoSAPIT™ reagent for every 5 µL of PCR product being cleaned
2. Use a thermal cycler to incubate at 37°C for 15 minutes followed by 80°C for 15 minutes
3. Store cleaned PCR products at –20°C until ready to sequence

Agarose Gel Preparation Protocol:

1. Create or obtain 1X TBE Electrophoresis Buffer. To create buffer from a solution, simply dilute the TBE solution to 1X using deionized water
2. Create a 1% agarose solution by combining 0.5 g of agarose with 50 mL of 1X TBE Buffer
3. Gently swirl the agarose solution and microwave the solution for 1 minute. Use autoclave gloves to remove the beaker from the microwave and gently swirl
4. Continue microwaving in intervals of 15-30 seconds until the agarose is completely dissolved and the solution is clear
5. Use a thermometer to monitor the solution as it cools to 65°C
6. When the solution is cooled to 65°C, add 2.5 μL of 10,000X Gel Red Gel Stain and swirl to mix
7. Pour the solution into a prepared gel tray with a gel comb inserted on the end closest to the edge.
8. Let the gel set for 20-30 minutes. The gel is set when it appears slightly opaque and does not move when the tray is jiggled

Preparing the Samples

1. Once the gel is poured, you can begin preparing the samples to be run
2. Prepare the samples by combining $1/10^{th}$ of the amplified product with 1 μL of loading buffer and distilled water to a total of 10 μL
3. Store the prepared samples in the fridge until ready to load into the gel

Running the Gel

1. Carefully remove the gel comb by pulling straight up.
2. Pour the 1X TBE Buffer into the anode (black/negative) side of the gel chamber first until the buffer reaches the gel
3. Pour the 1X TBE Buffer into the cathode (red/positive) side of the gel chamber next until the chamber is filled
4. Wait ~15 minutes, then pour the 1X TBE Buffer over the gel so it is 3-5 mm above the gel
5. Remove the prepared samples from the fridge and vortex each again to ensure they are well-mixed. Load 10 μL of 0.1X 100 bp ladder in the first well Load the amplified samples in the remaining wells
6. After loading all samples, place the cover on the gel chamber so the red and black wires are on the appropriate sides, and plug the wires into the matching sections of the power source
7. Turn on the power source and set it to 150 V for 30 minutes. Verify there are small bubbles on forming on both ends of the chamber. Check the location of the loading dye after 30 minutes and add more time if necessary

Sanger Sequencing Protocol:

Store all reagents on ice while preparing reactions

1. Add 4 to the number of samples to be sequenced to ensure enough master mix is made. This number will be N for the following step
2. Create the master mix by vortexing each of the following components and mixing the appropriate amount of each in a tube:
    a. N x 2.0 µL BigDye™ Terminator v3.1 Ready Reaction Mix
    b. N x 1.0 µL Sequencing Buffer
    c. N x 0.5 µL 10 µM primer
    d. N x 3.5 µL nuclease-free water
3. Pulse vortex the master mix for ten seconds and briefly centrifuge to remove any bubbles
4. Dispense 7.0 µL of master mix into each PCR tube
5. Vortex and briefly centrifuge PCR products. Add 3.0 µL of PCR product to each reaction
6. Pulse vortex each reaction for 10 seconds and centrifuge for 20-30 seconds.
7. Ensure there are no bubbles and perform a volume check
8. Load the reactions on a thermal cycler and run with the following parameters:

Hold	Cycling (25 cycles)			Hold
96°C	96°	50°C	60°C	12°C
1 min.	10 sec.	5 sec.	4 min.	∞

9. Store products at –20°C until ready for product cleanup

Sanger Sequencing Product Cleanup

4. Remove sequencing products from freezer and thaw. Keep on ice while adding the cleaning reagents.
5. Dispense 45 µL of SAM Solution to each product
6. Vortex the XTerminator Solution bulk container
7. Dispense 10 µL of XTerminator Solution to each product. Vortex XTerminator Solution bulk container before pipetting each time
8. If using a plate, seal the plate with heat seal or Clear Adhesive Film. If using tubes, secure the lids
9. Vortex products for 30 minutes
10. Centrifuge products briefly
11. Cleaned products can be stored at –20°C for up to two weeks

When ready to load products for capillary electrophoresis, only the clear upper phase should be used. If samples were cleaned on a plate, the BDx run module must be used. If samples were cleaned in tubes, 10 µL of the clear upper phase should be pipetted into a plate.

Sequencher® Data Upload and Analysis Protocol:

1. Open the Sequencher® software and drag samples into the window to upload sequencing data
2. To import the rCRS, navigate to File>Import>From Template>rCRS
   NOTE: This will automatically set the rCRS as a reference sequence with the appropriate base positions labeled to match the mtDNA genome. This rCRS only covers the control region
3. To align samples, select the desired samples including the rCRS and click "Assemble Automatically"
4. Rename the new contig appropriately and double click on the contig to see the information for the aligned sequences
5. To view the chromatograms, click "Bases" then "Show Chromatograms"
6. The top window shows the sequence of each aligned sample with the consensus sequence underneath. The bottom window shows the chromatograms for all samples with data at the selected consensus position
7. To find positions with disagreements in the base call between the aligned samples, click Command+D. This function is helpful for finding polymorphisms from the rCRS as well as any differences in the aligned sequences
8. To view a table of all disagreements between the aligned sequences and the reference, right-click the contig on the main Sequencher® window and select "Compare Bases to" then select "Reference Sequence". You can also choose to compare the bases to the consensus sequence to see all disagreements between all aligned sequences
9. A detailed report on any variances between aligned sequences and the rCRS can be viewed by selecting "Reports" then clicking "Entire Table" and selecting "Variance Detail Report" as the report format. This report provides snippets of the chromatograms at each position of disagreement as well as defines the primary and secondary peaks if an ambiguous base is called
10. Individual sequences or consensus sequences can be exported in various formats by right-clicking the sample/contig on the main Sequencher® window and selecting "Export" followed by the type of sequence you would like exported. The format can be changed using the drop-down arrow

6. Appendix B: Figures and Tables

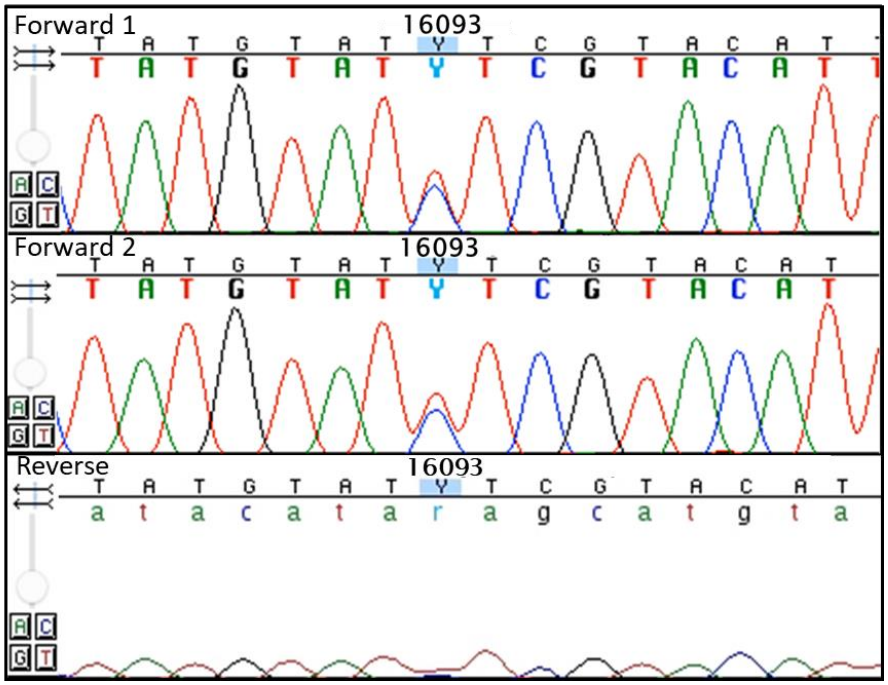

Figure 6.1: Chromatograms for Participant 3 buccal swab point heteroplasmy. The top two chromatograms show the light strand from two different primer reactions with overlapping sequences. The bottom chromatogram shows the heavy strand with position 16093 highlighted.

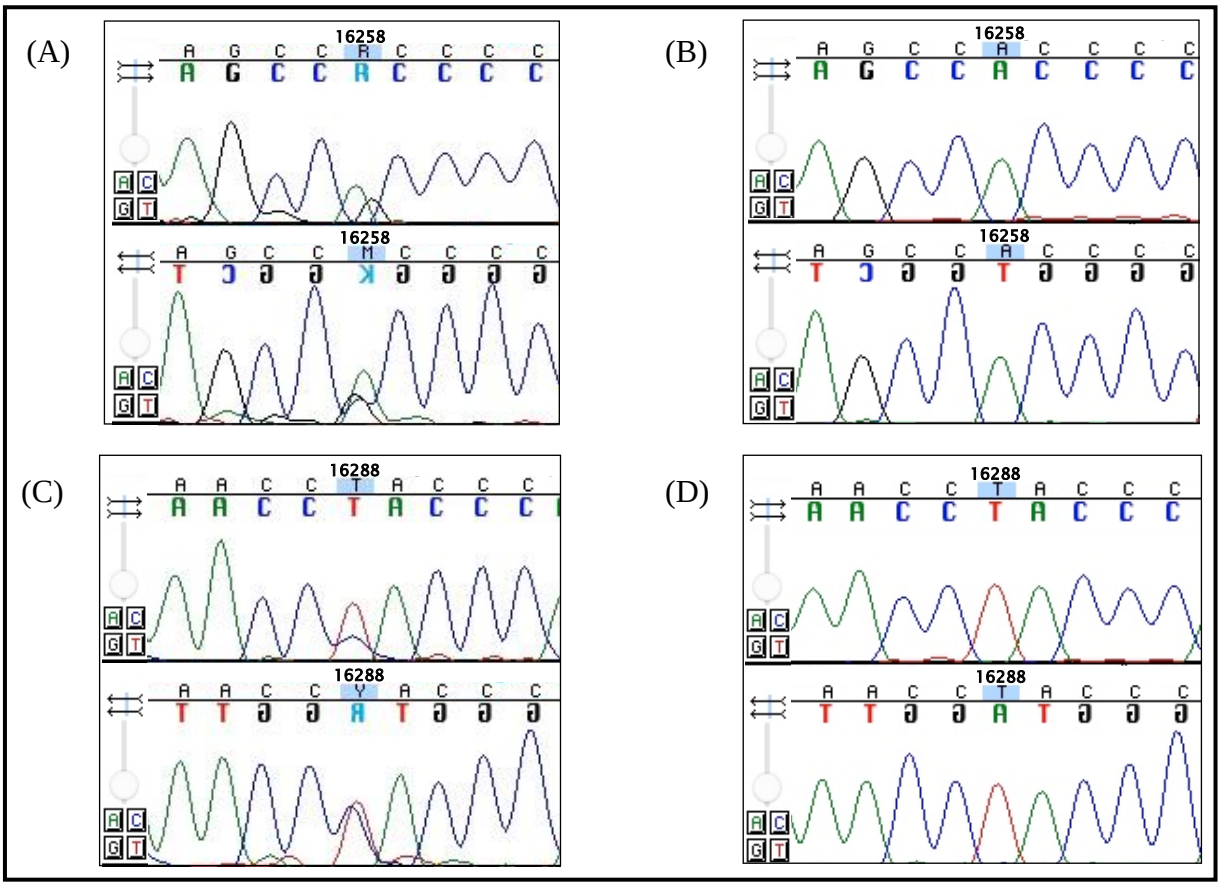

Figure 6.2: Chromatograms for Participant 3 hair shaft point heteroplasmy. For each set, the light strand is above the heavy strand. (A) Participant 3 hair shaft point heteroplasmy at position 16258. (B) Participant 3 reference buccal swab at position 16258. (C) Participant 3 hair shaft point heteroplasmy at position 16288. (D) Participant 3 reference buccal swab at position 16288.

Table 6.1: All sequence differences in reference buccal swabs from each participant.

	Position																					
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	7	1	1	1	1	2	3	4	4
	6	6	6	6	6	6	6	6	6	6	6	6	6	6	3	4	5	8	9	0	3	9	9
	0	2	2	2	2	2	2	2	3	3	3	3	3		6	2	9	5	0	1	7	9	
	0	2	2	4	7	8	9	9	0	1	2	5	9										
Participant	3	9	3	4	0	8	6	2	4	9	1	5	6	0									
1	T	T	T	T	A	T	T	C	T	G	T	T	T	A	G	C	C	A	C	A	A	C	G
2	C	C	T	T	G	C	C	T	C	A	C	T	T	G	G	T	C	G	T	G	G	C	G
3	T/C	T	C	C	A	C	C	C	C	A	C	T	T	G	G	T	T	A	T	A	A	T	G
4	T	T	C	T	A	C	C	C	C	A	T	C	T	G	A	T	T	A	T	A	A	C	G
5	T	T	C	T	A	C	C	C	C	A	T	T	C	G	G	T	T	A	C	A	A	C	A
rCRS*	T	T	C	T	A	C	C	C	C	A	T	T	T	G	A	T	T	A	T	A	A	C	G

*rCRS = revised Cambridge Reference Sequence

References

[1] Gill P, Jeffreys AJ, Werrett DJ. Forensic application of DNA ‘fingerprints.’ Nature. 1985;318(6046):577–579. doi:10.1038/318577a0

[2] Butler JM. Fundamentals of Forensic DNA Typing. Elsevier, Inc.; 2010.

[3] Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensics and mitochondrial DNA: applications, debates, and foundations. Annual Review of Genomics and Human Genetics. 2003;4:119–141. doi:10.1146/annurev.genom.4.070802.110352

[4] Butler JM, Levin BC. Forensic applications of mitochondrial DNA. Trends in Biotechnology. 1998;16(4):158–162. doi:10.1016/S0167-7799(98)01173-1

[5] Davis CL. Mitochondrial DNA: State of Tennessee v. Paul Ware. 1998. https://www.promega.com/-/media/files/resources/profiles-in-dna/103/mitochondrial-dna-state-of-tennessee-v-paul-ware.pdf?la=en#:~:text=Mitochondrial%20DNA%20(mtDNA)%20evidence%20was,four%2D%20year%2Dold%20girl.

[6] Graffy EA, Foran DR. A simplified method for mitochondrial DNA extraction from head hair shafts. Journal of Forensic Sciences. 2005;50(5):1119–1122.

[7] Lyons EA, Scheible MK, Sturk-Andreaggi K, Irwin JA, Just RS. A high-throughput Sanger strategy for human mitochondrial genome sequencing. BMC Genomics. 2013;14(1):881. doi:10.1186/1471-2164-14-881

[8] Tissue and Hair Extraction Kit (for use with DNA IQ). 2016.

[9] REPLI-g Mitochondrial DNA Kit Quick-Start Protocol. 2016.

[10] Human Mitochondrial DNA (mtDNA) Monitoring Primer Set. https://www.takarabio.com/documents/User%20Manual/7246_e.v1903Da-a_117027.pdf

[11] Gallimore JM, McElhoe JA, Holland MM. Assessing heteroplasmic variant drift in the mtDNA control region of human hairs using an MPS approach. Forensic Science International. Genetics. 2018;32:7–17. doi:10.1016/j.fsigen.2017.09.013

[12] Huang T. Next Generation Sequencing to Characterize Mitochondrial Genomic DNA Heteroplasmy. Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]. 2011;0 19:Unit19.8. doi:10.1002/0471142905.hg1908s71

[13] Tsiatis AC, Norris-Kirby A, Rich RG, Hafez MJ, Gocke CD, Eshleman JR, Murphy KM. Comparison of Sanger Sequencing, Pyrosequencing, and Melting Curve Analysis for the Detection of KRAS Mutations. The Journal of Molecular Diagnostics : JMD. 2010;12(4):425–432. doi:10.2353/jmoldx.2010.090188

[14] CODIS and NDIS Fact Sheet. Federal Bureau of Investigation. [accessed 2022 Mar 4]. https://www.fbi.gov/services/laboratory/biometric-analysis/codis/codis-and-ndis-fact-sheet

[15] Alberts B, Bray D, Hopkin K, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Essential Cell Biology. 4th ed. Garland Science; 2014.

[16] Speicher M, Antonarakis SE, Motulsky AG. Vogel and Motulsky's Human Genetics: Problems and Approaches. Springer Science & Business Media; 2009.

[17] Kivisild T. Maternal ancestry and population history from whole mitochondrial genomes. Investigative Genetics. 2015;6:3. doi:10.1186/s13323-015-0022-2

[18] Bergtrom G. Cell and Molecular Biology: What We Know & How We Found Out. 4th ed. Biological Sciences Faculty Books; 2020. https://dc.uwm.edu/biosci_facbooks/1/

[19] Genetics Glossary - ISOGG Wiki. [accessed 2022 Feb 15]. https://isogg.org/wiki/Genetics_Glossary

[20] van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Human Mutation. 2009;30(2):E386–E394. doi:10.1002/humu.20921

[21] Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. :6.

[22] DNA Maternal Ancestry Test – genebase. [accessed 2022 Feb 15]. https://www.genebase.com/product/dna-maternal-ancestry-test/

[23] DNA Testing for Ancestry & Genealogy | FamilyTreeDNA. [accessed 2022 Feb 15]. https://www.familytreedna.com/?cjevent=a1ad54d48e8011ec80f400910a82b839&cm_mmc=CJ-_-9166945-_-5292344-_-FamilyTreeDNA&utm_source=cj&utm_content=9166945&utm_term=13806157&cjdata=MXxOfDB8WXww

[24] Faro VL, Faro VL, Nolte IM, Brink JBT, Snieder H, Jansonius NM, Bergen AA, Bergen AA, Study LC. Mitochondrial Genome Study Identifies Association Between Primary Open-Angle Glaucoma and Variants in MT-CYB, MT-ND4 Genes and Haplogroups. Frontiers in Genetics. 2021 [accessed 2022 Feb 15];12. https://www.frontiersin.org/articles/10.3389/fgene.2021.781189/full. doi:10.3389/fgene.2021.781189

[25] Gusdon AM, Hui Y, Chen J, Mathews CE, Qu S. Mitochondrial haplogroup G is associated with nonalcoholic fatty liver disease, while haplogroup A mitigates the effects of PNPLA3. Endocrinology, Diabetes & Metabolism. 2021;4(1):e00187. doi:10.1002/edm2.187

[26] Kivisild T. English: mtDNA haplogroup tree and distribution map. Haplogroup labels are reported according to the http://www.phylotree.org/ nomenclature (van Oven M (2009)). Only a single branch defining marker, preferably from the coding region, is shown. The main

geographic features of haplogroup distribution are highlighted with colour. 2015. https://commons.wikimedia.org/wiki/File:MtDNA_haplogroup_tree_and_distribution_map.gif

[27] Butler JM. Advanced Topics in Forensic DNA Typing: Methodology. San Diego, California: Elsevier, Inc.; 2012.

[28] Sultana GNN, Sultan MZ. Mitochondrial DNA and Methods for Forensic Identification. Journal of Forensic Sciences & Criminal Investigation. 2018;9(1):1–5. doi:10.19080/JFSCI.2018.09.555755

[29] Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature Genetics. 1999;23(2):147–147. doi:10.1038/13779

[30] SWGDAM Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories. 2019. https://www.swgdam.org/_files/ugd/4344b0_f61de6abf3b94c52b28139bff600ae98.pdf

[31] Greenberg BD, Newbold JE, Sugino A. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene. 1983;21(1–2):33–49. doi:10.1016/0378-1119(83)90145-2

[32] Samuels DC, Li C, Li B, Song Z, Torstenson E, Boyd Clay H, Rokas A, Thornton-Wells TA, Moore JH, Hughes TM, et al. Recurrent Tissue-Specific mtDNA Mutations Are Common in Humans. PLoS Genetics. 2013;9(11):e1003929. doi:10.1371/journal.pgen.1003929

[33] Rebolledo-Jaramillo B, Su MS-W, Stoler N, McElhoe JA, Dickins B, Blankenberg D, Korneliussen TS, Chiaromonte F, Nielsen R, Holland MM, et al. Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(43):15474–15479. doi:10.1073/pnas.1409328111

[34] Li M, Rothwell R, Vermaat M, Wachsmuth M, Schröder R, Laros JFJ, van Oven M, de Bakker PIW, Bovenberg JA, van Duijn CM, et al. Transmission of human mtDNA heteroplasmy in the Genome of the Netherlands families: support for a variable-size bottleneck. Genome Research. 2016;26(4):417–426. doi:10.1101/gr.203216.115

[35] Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. Nature Genetics. 1996;12(4):417–420. doi:10.1038/ng0496-417

[36] Zhang H, Burr SP, Chinnery PF. The mitochondrial DNA genetic bottleneck: inheritance and beyond. Essays in Biochemistry. 2018;62(3):225–234. doi:10.1042/EBC20170096

[37] Zaidi AA, Wilton PR, Su MS-W, Paul IM, Arbeithuber B, Anthony K, Nekrutenko A, Nielsen R, Makova KD. Bottleneck and selection in the germline and maternal age influence transmission

of mitochondrial DNA in human pedigrees. Proceedings of the National Academy of Sciences. 2019;116(50):25172–25178. doi:10.1073/pnas.1906331116

[38] Salas A, Lareu MV, Carracedo A. Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: a case report. International Journal of Legal Medicine. 2001;114(3):186–190. doi:10.1007/s004140000164

[39] Li M, Schönberg A, Schaefer M, Schroeder R, Nasidze I, Stoneking M. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. American Journal of Human Genetics. 2010;87(2):237–249. doi:10.1016/j.ajhg.2010.07.014

[40] Paneto GG, Longo LVG, Martins JA, de Camargo MA, Costa JC, de Mello ACO, Chen B, Oliveira RN, Hirata MH, Cicarelli RMB. Heteroplasmy in hair: study of mitochondrial DNA third hypervariable region in hair and blood samples. Journal of Forensic Sciences. 2010;55(3):715–718. doi:10.1111/j.1556-4029.2010.01339.x

[41] Rasmussen EM, Sørensen E, Eriksen B, Larsen HJ, Morling N. Sequencing strategy of mitochondrial HV1 and HV2 DNA with length heteroplasmy. Forensic Science International. 2002;129(3):209–213. doi:10.1016/s0379-0738(02)00276-1

[42] Irwin J, Saunier J, Niederstätter H, Strouss K, Andreaggi K, Diegoli T, Kloss-Brandstätter A, Parson W, Parsons T. Investigation of Heteroplasmy in the Human Mitochondrial DNA Control Region: A Synthesis of Observations from More Than 5000 Global Population Samples. Journal of molecular evolution. 2009;68:516–27. doi:10.1007/s00239-009-9227-4

[43] Grady JP, Pickett SJ, Ng YS, Alston CL, Blakely EL, Hardy SA, Feeney CL, Bright AA, Schaefer AM, Gorman GS, et al. mtDNA heteroplasmy level and copy number indicate disease burden in m.3243A>G mitochondrial disease. EMBO molecular medicine. 2018;10(6):e8262. doi:10.15252/emmm.201708262

[44] Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. Nature Reviews Genetics. 2015;16(9):530–542. doi:10.1038/nrg3966

[45] Paneto GG, Martins JA, Longo LVG, Pereira GA, Freschi A, Alvarenga VLS, Chen B, Oliveira RN, Hirata MH, Cicarelli RMB. Heteroplasmy in hair: Differences among hair and blood from the same individuals are still a matter of debate. Forensic Science International. 2007;173(2):117–121. doi:10.1016/j.forsciint.2007.02.011

[46] Sekiguchi K, Sato H, Kasai K. Mitochondrial DNA heteroplasmy among hairs from single individuals. Journal of Forensic Sciences. 2004;49(5):986–991.

[47] Roberts KA, Calloway C. Characterization of Mitochondrial DNA Sequence Heteroplasmy in Blood Tissue and Hair as a Function of Hair Morphology*,†,‡. Journal of Forensic Sciences. 2011;56(1):46–60. doi:10.1111/j.1556-4029.2010.01540.x

[48] Barrett A, Arbeithuber B, Zaidi A, Wilton P, Paul IM, Nielsen R, Makova KD. Pronounced somatic bottleneck in mitochondrial DNA of human hair. Philosophical Transactions of the Royal Society B: Biological Sciences. 2020;375(1790):20190175. doi:10.1098/rstb.2019.0175

[49] Colon NS. VARIATION IN MITOCHONDRIAL DNA HETEROPLASMY FROM BLOOD, BUCCAL, AND HAIR SAMPLES. Fort Worth, TX: University of North Texas Health Science Center; 2020.

[50] Allen M, Engström A-S, Meyers S, Handt O, Saldeen T, von Haeseler A, Pääbo S, Gyllensten U. Mitochondrial DNA Sequencing of Shed Hairs and Saliva on Robbery Caps: Sensitivity and Matching Probabilities. Journal of Forensic Sciences. 1998;43(3):16169J. doi:10.1520/JFS16169J

[51] Grzybowski T. Extremely high levels of human mitochondrial DNA heteroplasmy in single hair roots. Electrophoresis. 2000;21(3):548–553. doi:10.1002/(SICI)1522-2683(20000201)21:3<548::AID-ELPS548>3.0.CO;2-U

[52] Budowle B, Allard MW, Wilson MR. Critique of interpretation of high levels of heteroplasmy in the human mitochondrial DNA hypervariable region I from hair. Forensic Science International. 2002;126(1):30–33. doi:10.1016/S0379-0738(02)00019-1

[53] Quality Assurance Standards for Forensic DNA Testing Laboratories. 2020. https://www.fbi.gov/file-repository/quality-assurance-standards-for-forensic-dna-testing-laboratories.pdf/view

[54] Wolfram LJ. Reactivity of Disulphide Bonds in Strained Keratin. Nature. 1965;206(4981):304–305. doi:10.1038/206304a0

[55] DNA IQ System -- Small Sample Casework Protocol. 2021.

[56] QIAamp DNA Investigator Kit Handbook. 2020.

[57] Kretz K, Callen W, Hedden V. Cycle sequencing. Genome Research. 1994;3(5):S107–S112.

[58] Sanger Sequencing Steps and Method. Millipore Sigma.

[59] Chaitanya L, van Oven M, Brauer S, Zimmermann B, Huber G, Xavier C, Parson W, de Knijff P, Kayser M. High-quality mtDNA control region sequences from 680 individuals sampled across the Netherlands to establish a national forensic mtDNA reference database. Forensic Science International. Genetics. 2016;21:158–167. doi:10.1016/j.fsigen.2015.12.002

[60] FoswikiMITOMASTER. https://www.mitomap.org/foswiki/bin/view/MITOMASTER/WebHome

[61] Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, et al. The Molecular Dissection of mtDNA Haplogroup H Confirms That the Franco-Cantabrian Glacial Refuge Was a Major Source for the European Gene Pool. The American Journal of Human Genetics. 2004;75(5):910–918. doi:10.1086/425590

[62] Mita Y, Fukagawa T, Watahiki H, Kitayama T, Fujii K, Mizuno N, Sekiguchi K. Developmental validation for Sanger sequencing of HV1 and HV2 in mitochondrial DNA. Forensic Science International: Reports. 2020;2:100159. doi:10.1016/j.fsir.2020.100159

[63] Gabriel MN, Huffine EF, Ryan JH, Holland MM, Parsons TJ. Improved MtDNA sequence analysis of forensic remains using a “mini-primer set” amplification strategy. Journal of Forensic Sciences. 2001;46(2):247–253.

[64] Desmyter S, Bodner, Huber G, Dognaux S, Berger C, Noël F, Parson W. Hairy matters: MtDNA quantity and sequence variation along and among human head hairs - ClinicalKey. 2016;25:1–9.

[65] Sekiguchi K, Imaizumi K, Matsuda H, Mizuno N, Yoshida K, Senju H, Sato H, Kasai K. MtDNA Sequence Analysis Using Capillary Electrophoresis and Its Application to the Analysis of MtDNA in Hair. Japanese journal of science and technology for identification. 2003;7(2):123–130. doi:10.3408/jasti.7.123

[66] Andréasson H, Nilsson M, Budowle B, Lundberg H, Allen M. Nuclear and mitochondrial DNA quantification of various forensic materials. Forensic Science International. 2006;164(1):56–64. doi:10.1016/j.forsciint.2005.11.024

[67] Brandhagen MD, Loreille O, Irwin JA. Fragmented Nuclear DNA Is the Predominant Genetic Material in Human Hair Shafts. Genes. 2018;9(12):640. doi:10.3390/genes9120640

[68] Melton T, Dimick G, Higgins B, Lindstrom L, Nelson K. Forensic Mitochondrial DNA Analysis of 691 Casework Hairs. 2005 [accessed 2022 Feb 24];50(1). https://www.astm.org/jfs2004230.html

[69] Roberts KA, Calloway C. Mitochondrial DNA amplification success rate as a function of hair morphology. Journal of Forensic Sciences. 2007;52(1):40–47. doi:10.1111/j.1556-4029.2006.00322.x

[70] Eckhart L, Bach J, Ban J, Tschachler E. Melanin binds reversibly to thermostable DNA polymerase and inhibits its activity. Biochemical and Biophysical Research Communications. 2000;271(3):726–730. doi:10.1006/bbrc.2000.2716

[71] Opel KL, Chung D, McCord BR. A study of PCR inhibition mechanisms using real time PCR. Journal of Forensic Sciences. 2010;55(1):25–33. doi:10.1111/j.1556-4029.2009.01245.x

[72] Damaso N, Ashe EC, Meiklejohn KA, Kavlick MF, Robertson JM. Comparison of polymerases used for amplification of mitochondrial DNA from challenging hairs and hairs of various treatments. Forensic Science International: Genetics. 2021;52:102484. doi:10.1016/j.fsigen.2021.102484

[73] Forster L, Forster P, Lutz-Bonengel S, Willkomm H, Brinkmann B. Natural radioactivity and human mitochondrial DNA mutations. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(21):13950–13954. doi:10.1073/pnas.202400499

[74] Kim BM, Hong SR, Chun H, Kim S, Shin K-J. Comparison of whole mitochondrial genome variants between hair shafts and reference samples using massively parallel sequencing. International Journal of Legal Medicine. 2020;134(3):853–861. doi:10.1007/s00414-019-02205-y

[75] Tully LA, Parsons TJ, Steighner RJ, Holland MM, Marino MA, Prenger VL. A sensitive denaturing gradient-Gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. American Journal of Human Genetics. 2000;67(2):432–443. doi:10.1086/302996

[76] Budowle B, DiZinno JA, Wilson MR. INTERPRETATION GUIDELINES FOR MITOCHONDRIAL DNA SEQUENCING. :9.

[77] Zhao H, Shen J, Medico L, Platek M, Ambrosone CB. Length heteroplasmies in human mitochondrial DNA control regions and breast cancer risk. International Journal of Molecular Epidemiology and Genetics. 2010;1(3):184–192.

[78] Berger C, Hatzer-Grubwieser P, Hohoff C, Parson W. Evaluating sequence-derived mtDNA length heteroplasmy by amplicon size analysis. Forensic Science International. Genetics. 2011;5(2):142–145. doi:10.1016/j.fsigen.2010.10.002

[79] SWGDAM Contamination Prevention and Detection Guidelines for Forensic DNA Laboratories. 2017.

[80] Date-Chong M, Buoncristiani MR, Aceves M, Orrego C. An examination of the utility of a nuclear DNA/mitochondrial DNA duplex qPCR assay to assess surface decontamination of hair. Forensic Science International. Genetics. 2013;7(3):392–396. doi:10.1016/j.fsigen.2013.03.003

[81] Naue J, Sänger T, Lutz-Bonengel S. Get it off, but keep it: Efficient cleaning of hair shafts with parallel DNA extraction of the surface stain - ClinicalKey. 2020 [accessed 2022 Mar 30];45. https://www.clinicalkey.com/#!/content/playContent/1-s2.0-S1872497319304272?returnurl=null&referrer=null

[82] Cline J, Braman JC, Hogrefe HH. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Research. 1996;24(18):3546–3551. doi:10.1093/nar/24.18.3546

[83] Shinde D, Lai Y, Sun F, Arnheim N. Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. Nucleic Acids Research. 2003;31(3):974–980. doi:10.1093/nar/gkg178

[84] Nakai T, Sakurada A, Endo T, Kobayashi H, Masuda S, Makishima M, Esumi M. Caution for simple sequence repeat number variation in the mitochondrial DNA D-loop to determine cancer-specific variants. Oncology Letters. 2019;17(2):1883–1888. doi:10.3892/ol.2018.9809