THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

FORENSIC SCIENCE PROGRAM

INVESTIGATING RATES OF HETEROPLASMY IN THE MITOCHONDRIAL DNA CONTROL REIGON OF AFRICAN, ASIAN, AND LATINO POPULATION GROUPS USING MASSIVELY PARALLEL SEQUENCING

EMILY DEMCHAK
SUMMER 2018

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Forensic Science
with honors in Forensic Science

Reviewed and approved* by the following:

Mitchell M. Holland
Associate Professor, Biochemistry & Molecular Biology
Thesis Supervisor

Jennifer A. McElhoe
Assistant Research Professor, Biochemistry & Molecular Biology
Faculty Reader

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

Massively parallel sequencing (MPS), a high-throughput form of next generation sequencing, allows increased resolution of mitochondrial (mt) DNA heteroplasmy and is at the forefront of efforts to expand the utility of forensic mtDNA typing. Heteroplasmy is a heterogeneous collection of sequence variants in the cytoplasm of the cell. It is hypothesized that there is potential for differences in rates of heteroplasmy linked to population haplogroups, based on assumption and empirical observation that the position and rate of heteroplasmy may be linked to the haplotype sequence. The current project has used an MPS approach to measure, analyze, and report rates of heteroplasmy on per sample and per nucleotide basis for 377 samples in population groups reporting to be non-European (NIJ-2016-DN-BX-0171). Buccal cells were collected from unrelated non-European individuals and MPS analysis conducted on the control region (CR) of the mtDNA genome using Nextera ® XT library preparation and 150X150 paired-end reads on an Illumina MiSeq. Secondary analysis was performed using GeneMarker HTS software to evaluate haplotype and heteroplasmy, and HaploGrep2 to determine haplogroups. Heteroplasmy was shown to occur in the African population at a rate of 33%, the Asian population at 31.4%, and the Latino population at 26.4%. If heteroplasmy did occur, it was more likely to be at a minor variant percentage below 10% and one site of heteroplasmy within an individual was the most common occurrence. Position 16093 is a consistent hot-spot for heteroplasmy across all populations, with other hot spots at homoploymeric regions and in the range of positions 185-215. This thesis shows partial results of a larger study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to specially thank Dr. Mitchell Holland and Dr. Jennifer McElhoe for their amazing guidance, education, support, and perseverance through the progress of this project. Thank you for the priceless experience of training under both of you. Thank you to the National Institute of Justice for providing funding for this project (NIJ 2016-DN-BX-0171).

Thank you to my parents; Dad, Rhonda, and Mom, for your belief in me, your encouragement and support, and your pride in my accomplishments. Thank you to all of my friends and family, especially my research partner Maddie Crowley for contributing data and for spending so much time in Thomas basement with me that our GPS began to think we lived there.

Thank you to the Schreyer Honors College for the opportunity to pursue this project and graduate with esteemed honors. Thank you for the financial support of a research grant to allow the summertime work and for the financial support to present this project at the American Academy of Forensic Sciences 70th annual meeting.

Thank you to the Forensic Science program full of faculty, administration, and students who are as kind as they are intelligent, interesting, supportive, and full of valuable experience. Also, thank you for the financial support to present this research at the North Eastern Association of Forensic Scientist's 43rd annual meeting and American Academy of Forensic Sciences 70th annual meeting

Lastly, thank you to my wonderful partner Skyler. You have been here through this whole ride. Thank you for always being a great listener and comforter. Thank you for always pushing me toward the finish and trying to keep my thoughts positive. Thank you for prioritizing my health as well as my goals. I'm incredibly grateful to have you by my side, I love you.

Introduction

DNA profiling is seen as the most robust form of identification available to a forensic investigation. Typical DNA analysis is performed by examining markers along the human genome to reveal the unique profile of an individual. However, not all cases provide the biological material necessary for this analysis. Mitochondrial DNA (mtDNA) testing supplements the choices available to forensic analysis. Samples with little or no chromosomal DNA, such as hair shafts and old bone specimens, are most often the targets for mtDNA analysis (Melton & Nelson, 2001). Although mtDNA profiles are not unique to an individual, matches provide strong circumstantial information in a forensic investigation. Therefore, it would benefit the forensic community to increase the discrimination power of mtDNA testing.

Human mitochondria possess a circular extrachromosomal genome consisting of 16,569 base pairs (bps) (Anderson et al, 1981) that is inherited maternally. The mitochondrial genome (mtgenome) encodes for 37 genes and contains a non-coding region of approximately 1,122 bp referred to as the control region (CR) (Holland & Parsons, 1999). The CR spans nucleotide positions 16,024 to 16,569 and 1 to 576 and contains hypervariable region 1 (HV1; positions 16,024 to 16,365) and hypervariable region 2 (HV2; positions 73 to 340). While the hypervariable regions have historically been the targets of forensic mtDNA analysis (Holland & Parsons, 1999; SWGDAM, 2013), the recent development of massively parallel sequencing (MPS), a form of next generation sequencing (NGS), has allowed for easier analysis of the CR and the entire mtgenome (Gallimore et al, 2018; Holland et al, 2018; Chaitanya et al, 2015; Chaitanya et al, 2016).

The oocyte is the source of mitochondria for the offspring embryo during fertilization, therefore mtDNA is inherited maternally (Holland & Parsons, 1999). A maternal lineage will share the same collection of single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) called the haplotype. Often non-related individuals will present different haplotypes, however certain haplotypes are common in population groups and may be shared among unrelated individuals (Carracedo et al, 2000). In line with this, a haplogroup consists of similar haplotypes which have risen from related ancestral lineages. Due to mutational events, and the fact that there are hundreds to thousands of copies of mtDNA in most cells (Holland & Parsons, 1999), sequence variants may arise in lower proportions in addition to the primary SNPs and indels that make up the haplotype. This state is referred to as heteroplasmy; a heterogeneous collection of sequence variants in the cytoplasm of the cell. A variation at a nucleotide position (np) in the mtgenome may occur in any number of the mitochondria present in an individual, which allows for a potentially wide range in the proportion of variants present when a sample is sequenced and analyzed (Gallimore et al, 2018).

Reactive oxygen species present within the mitochondrion due to cellular respiration cause a majority of point heteroplasmy (PHP). Length heteroplasmy (LHP) occurs in the CR often in longer stretches of GC base pairs (Forester et al, 2010), where the γ-polymerase struggles to replicate homopolymeric stretches longer than 8 nucleotides. The replicative slippage which causes LHP is similar to that which occurs in polymerase chain reaction (PCR) where the polymerase may unattach and then realign slightly ‘behind’ the original position and cause extra bases to be inserted. It is believed that heteroplasmy occurs as a result of these replicative anomalies or as a transitional state where the nucleotide position is changing from one base to

another. Interestingly, the vast majority of these mutated copies of the mtgenome are selectively removed through a purifying process that occurs in the mitochondrion (Greaves et al, 2014), contributing to the relative rarity of observing heteroplasmic variants. The presence of heteroplasmic variants logically increases the identifying weight of an mtDNA sequence match.

Heteroplasmy has been used as compelling evidence in the identification of unknown biological samples. In practice, if the sequence variant is present as greater than 50% of the copies of mtDNA in the cell, then it is considered a haplotype SNP or indel. Any variants that do not reach the 50% level are considered heteroplasmic. The first and most well-known example of observing heteroplasmy in a forensic investigation was the identification of the skeletal remains of Tsar Nicholas Romanov (Gill et al., 1994, Ivanov et al, 1996). The latter study compared the putative mtDNA sequence of the Tsar to the mtDNA profile generated from the remains of the Tsar's brother, Georgji. The two DNA profiles were a haplotype sequence match and shared heteroplasmy at np 16169 of the CR; numbering according to Anderson et al, 1981. The weight of the mtDNA match was calculated through a likelihood ratio (LR) approach: the probability that the remains are the questioned sample divided by the probability that the remains are an unrelated sample. Ivanov et al. (1996) considered the haplotype and heteroplasmy as independent events which allowed the haplotype LR (150) to be multiplied by the heteroplasmy LR (2,500), resulting in a combined LR of 375,000. Explained, the likelihood that the samples tested were the Tsar's is 375,000 times greater than the likelihood that the samples tested were a random individual. This case illustrated that an mtDNA match can be powerful evidence, especially when heteroplasmy is involved.

Although heteroplasmy is typically too low to be reliably detected and reported with the conventional Sanger sequencing technique (Salas et al, 2001), the introduction of MPS has allowed

heteroplasmy to be reliably detected and resolved to a 2% reporting threshold as established in the Holland laboratory (Rathbun et al, 2017, Gallimore et al, 2018). The Holland laboratory has developed an MPS method for sequencing the CR and has previously analyzed 545 samples of mtDNA from individuals of European descent (NIJ 2014-DN-BX-K022). The current project is modeled after the European study to investigate rates of heteroplasmy among African, Asian and Latino populations and will compare newly generated data against the European dataset.

Point heteroplasmy (PHP) in the European study was assessed for samples analyzed at a threshold of 2%. PHP was observed in 41% of all samples (221 individuals), for a total of 283 heteroplasmic observations in 545 individuals. Most individuals had one site of heteroplasmy (174/545 or 32%) while 33 individuals had two sites (6%), thirteen had three sites (2.4%), and one individual had four locations of heteroplasmy (0.18%). The CR contains 1122 nps, however only a total of 80 (7.1%) of these sites exhibited at least one occurrence of heteroplasmy. Of the 283 observed instances of heteroplasmy, 111 (39.2%), 152 (53.7%), and 20 (7.1%) of the observations fell within HVI, HVII, and outside the hypervariable regions, respectively. The most common site of PHP, position 16093 (12.4% of total observations), is a common haplotype variant. The high occurrence of heteroplasmy at this position supports the assumption that heteroplasmy can function as a transitional state of variance within the mtgenome. Linages with a fixed haplotype variant at this position may have once had heteroplasmy here in their ancestors. Overall, these rates from the European study are evidence that heteroplasmy is a rare but significant factor in the CR of mtDNA, and analysis via MPS methods is improving their current understanding.

Samples for this study were collected from individuals who self-reported as Black, African, or African American; Hispanic or Latino; and East Asian or who reported their mother and/or maternal grandmother's ancestry as African, Asian, Native American, Middle Eastern, Indian, or

Hispanic. These samples were then sequenced and their haplogroups were determined using HaploGrep2, an online tool based on PhyloTree build 17 (Van Oven, 2015; Weissensteiner et al, 2016) The population groups of interest for this project contain samples sorted by haplogroups, known to belong to African, Asian, and Latino lineages. Haplogroups are named by a series of letters and numbers, with the first letter referring to the basal node of the phylogenetic branch on which that haplotype lies, and subsequent letters and numbers characterizing subgroups, with the group becoming more specific as characters are added.

It is hypothesized that there is potential for differences in rates of heteroplasmy for European, African, Asian, and Latino population groups. We have tested this hypothesis based on an assumption (and through empirical observation) that the position and rate of heteroplasmy may be linked to the haplotype sequence. The forensic science community requires further validation of MPS analysis of mtDNA to encourage adoption of this technology into working laboratories and investigations. Recent studies within the Holland laboratory have produced a reliable framework for mtDNA analysis, and preliminary results on the European population group will serve as a useful comparison against the questioned population groups. This project is an appropriate expansion for needed validation of MPS methods and will provide a deeper understanding of rates of heteroplasmy across various population groups.

Materials and Methods

All work for this study has been conducted under Penn State University internal review board (IRB) approved projects STUDY00000970 (Holland lab) and PRAMS00045727 (Shriver lab).

Sample Collection and DNA Extraction Procedures

A total of 789 samples were collected by the Holland group and the Shriver group. Buccal swabs collected by the Holland group (n=280) were extracted using a buccal brush purification protocol from Gentra Puregene® Buccal Cell Kit (Qiagen, 2011). Liquid saliva samples (n=321) provided by the Shriver group were extracted using a body fluid purification protocol from Gentra Purgene® Blood Kit (Qiagen, 2011), and 188 Shriver group collected buccal samples were extracted by the Penn State Genomics Core Facility (University Park, PA).

Control Region PCR Enrichment

A fragment of mtDNA 1149 basepairs (bps) in length containing the entire CR was amplified. Table 1 shows the two transposase adapted oligonucleotide primer sets which were used to amplify the CR using a PCR program with activation of the polymerase at 94 ºC for 1 min, followed by 35 cycles of 98 ºC for 10 sec, 56 ºC for 30 sec, 68 ºC for minute, 72 ºC for 5 min and a 4 ºC hold. The transposase sequence was included to ensure the capture of the outermost ends of the target sequence during library preparation and sequencing.

Table 1: PCR primer sequences.

Primer	Transposase Sequence (5'→3')*	Primer Sequence (5'→3')	Positions	Reference
F15997-A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	CACCATTAGCACCCAAAGCT	15978-15997	Illumina*
F15997-B	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	CACCATTAGCACCCAAAGCT	15978-15997	Illumina*
R590-A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	TCAGTGTATTGCTTTGAGGAGGT	590-612	this study
R590-B	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	TCAGTGTATTGCTTTGAGGAGGT	590-612	this study

Table courtesy of J. McElhoe, Ph.D.
*Primer set F15997 and transposase sequence taken from Illumina 16S Metagenomics Sequencing Library Preparation (Part # 15044223 Rev. B).

Library preparation

PCR products were Qubit (Invitrogen, Carlsbad, CA) quantified before library preparation was conducted using Nextera® XT kit reagents (Illumina, San Diego, CA). Library preparation consists of four steps: tagmentation, indexing, clean up, and loading.

Tagmentation included the fragmenting and tagging of input DNA. Proper fragment size is essential for adequate cluster generation on the flow cell to allow nucleotide signals to be read by the sequencer. Tagging allows for the addition of individualized index pairs for each sample, based on complimentary adaptor sequences. The utility of the transposase adapted PCR primers occurred with this step; the beginning and end of the target sequence was not lost during randomized tagmentation.

The indexing step follows tagmentation where individualized index pairs were added to the tagmented library. Indices were added like a checkerboard across a sample well plate, thus each individual well contains a unique pair of indices. This unique set of identifiers allowed

pooling of samples in one sequencing reaction, therefore greatly increasing the possible number of samples within one sequencing run. The index adaptors also contained primer sequences which are complimentary to forward and reverse primers attached to the glass flow cell, so this step allowed the libraries to attach and move forward with the sequencing chemistry.

PCR clean-up followed indexing which was a short PCR cycle. Agencourt® AMPure® XP beads (Beckman Coulter Inc., Indianapolis, IN) were used to remove impurities and excess tagmentation and index primers. AMPure® XP beads reversibly bind DNA in the presence of propylene glycol salt. These magnetic beads trapped the DNA and a specialized tray concentrated the beads in one area, which allowed the user to remove unwanted solution and wash the DNA products. A resuspension solution was then used to retrieve the DNA off the beads.

The final step of library preparation was normalization, denaturation, and loading onto the MiSeq sequencer. The first two runs of this report utilized Netera® XT normalization beads, which function similarly to the AMPure® XP beads, and were intended to retain a controlled amount of DNA and release it in denatured form. However, quantification of the normalization bead product showed bi-modal results which did not match theoretical normalization. Therefore, subsequent runs utilized Qubit quantification and manual normalization and denaturation. Samples of the target concentration of 2 nM were spiked with 1% PhiX internal control and loaded onto the MiSeqFGX with a final target concentration of 11-12 pM for adequate clustering on our machine.

MPS sequencing on the MiSeqFGx

All sequencing runs were performed on a MiSeq FGx benchtop sequencer (Illumina, San Diego, CA) using 150 x 150 (300 cycle kit; v.2 chemistry; Illumina Inc.) paired-end reads. The

tagmented and indexed samples attached to a primer lawn present on a glass flow cell, utilizing the adaptor sequences added in the indexing step. The library fragments then folded over and performed bridge amplification to form clusters. The reverse fragments were cleaved and washed away for the first set of 150 reads. Fluorescent dye tagged reversible terminating dNTPs were washed over the flow cell, and the corresponding base incorporated into a cluster would fluoresce following excitation, allowing an on-board charge-coupled device to read the base sequence of that cluster. The fluorescent dyes were then removed before the 3' OH group was reconstituted, allowing new dNTPs to be washed over the flow cell to become incorporated, and thus the reading steps were repeated. At the end of the forward reads, the fragments bridges amplify once more and then the forward strands are washed away. The second 150 nps were read from the reverse strand.

MPS Data Analysis

GeneMarker® HTS (Holland et al, 2017; SoftGenetics, State College, PA) was used for secondary analysis of the fastq files generated from MiSeq Reporter. Sequences were aligned to the revised Cambridge Reference Sequence (rCRS; GenBank ID NC_012920.1) (Andrews, et al. 1999) and a custom motif file was used to ensure phylogenetically correct calls. Reported haplotypes included SNPs that occurred in >50% of the reads. Potential point heteroplasmy (PHP) was considered for the minor variants that occurred ≤50% of the reads. The following filters were applied to generate reporting tables: threshold for minor variants of 2.0% or greater of the total reads; variant allele coverage of at least 40 reads, total coverage of at least 200 reads, an allele score difference ratio ≤10%, and allele balance ratio ≤2.5% for SNPs and ≤5.0% for indels. Therefore, the variant reports included variant sequences where each nucleotide position has a

coverage of at least 200, the variant was detected in at least forty reads, and minor variants were present in at least 2.0% of the total reads. This deep coverage MPS (DCMPS) allowed the improved resolution of heteroplasmic variants over traditional Sanger sequencing (Holland et al, 2018). Read threshold values were used to ensure reliable reporting of variants at the 2% level. A total of 2% of the total reads at any one position were required to report a minor variant. The minimum coverage threshold for these minor variants was 40 calls: 2% of 2000 reads. With the requirement of 40 variant calls, nucleotide positions with a depth of coverage less than 2000 calls cannot be resolved down to the 2% variant percentage. For example, a nucleotide position with 40 variant calls out of a total of 1500 calls produces a variant percentage of 2.6%. Due to this, low coverage was considered to be below 2000 total reads because this was where the minimum 40 variant calls began to push heteroplasmic resolution above the 2% level.

Figure 1: Example pileup image taken from GeneMarker® HTS showing the reference sequence, consensus sequence, and limited view of the pileup of aligned sequenced reads.

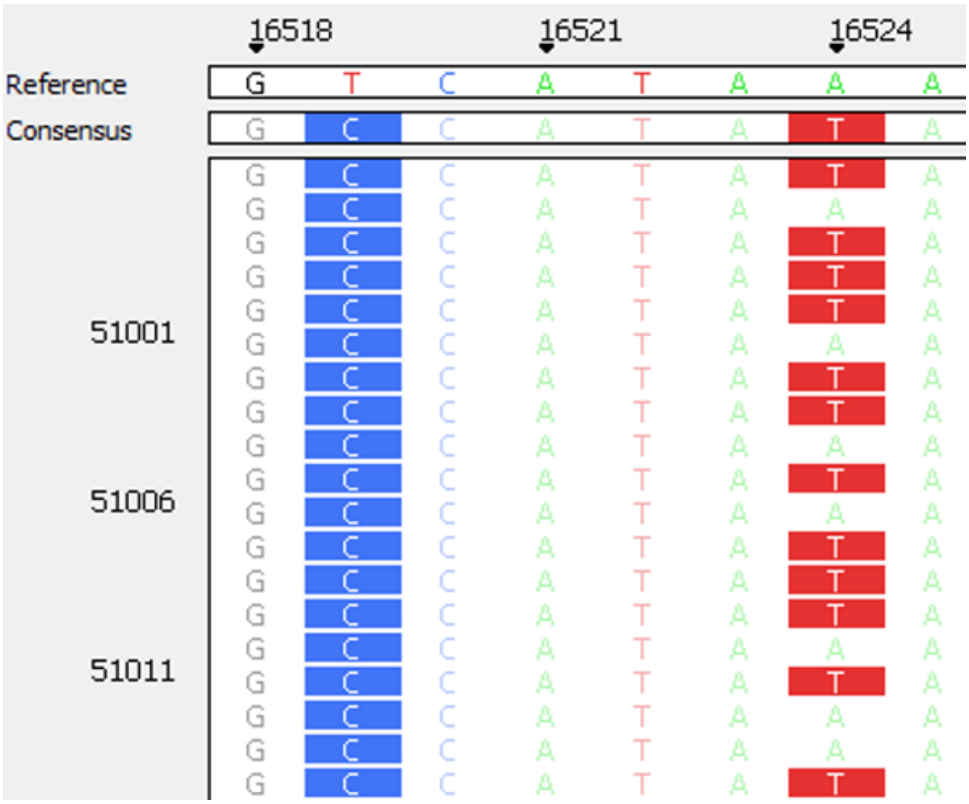


Figure 1 shows a SNP and PHP variant in the GeneMarker® HTS viewer. The top line is the reference rCRS sequence, with the next line being the consensus sequence, which represents the major sequence of the sample. Below these lines, the pileup shows individual reads where a SNP without heteroplasmy can be seen at position 16519 and heteroplasmy can be seen at position 16524 where the reference A intervenes periodically between major SNP T at this position.

Contamination of samples was determined by assessing variant reports and the GeneMarker® HTS pileup. Minor variants of all samples, especially those which showed 5 or more, were carefully scrutinized within the software to determine authenticity and if contamination was present. If there were uninterrupted read stretches in the pileup which contained multiple sites

with PHP and similar minor variant frequencies, then the sample was considered to be contaminated.

HaploGrep2 (Van Oven, 2015; Weissensteiner et al, 2016) was used to determine haplogroups. The African group is characterized by Haplogroup L, which is the oldest human mtDNA haplogroup and composes the base of the mtDNA phylogenetic tree (Chen et al., 1995; van Oven, 2015). The Asian group for this study is characterized by the haplogroups A (excluding the majority of A2, but including A2a and A2b), B4, B5, C1a, C4a, C4b, C5, D4, D5, E, F, G, M7, M8a, M9, M10, M11, M12, N9a, N9b, R9a, Y, and Z (Allard et al., 2004; Tanaka et al, 2004; Kivisild et al., 2002; Umetsu and Yuasa, 2005). These haplogroups and their subgroups are of Eastern Asian origin and do not include haplogroups which are considered Eurasian or Indian.

Due to the colonization of South America and the trans-Atlantic slave trade, modern Hispanic/Latino populations are known to be admixed; a combination of European, African, and Native American genetic lineages (Bryc et al, 2010). Mitochondrial DNA does not recombine like chromosomal DNA, so maternal ancestry is preserved in the process of recent admixture. Therefore; a modern person identifying as Latino can have a maternal haplotype which falls into European, African, or Native American haplogroups, and this lineage may be well beyond the genealogical knowledge of the individual. This study characterizes the Latino population group as the Native American haplogroups A2 (excluding A2a and A2b), B2, C1b, C1c, C1d, C4c, D1, D4h3a, and D4e1c (Alves-Silva et al., 2000; Kumar et al., 2011; Perego et al., 2010). These haplogroups are considered to be Native Latin American or Pan-American and common in Latin countries; the North American X2a is not included as Latino (Perego et al., 2009).

Results

Six hundred thirty-five (635) samples were sequenced for this report. Of these samples, 377 were included in the final data analysis for the thesis exercise. As table 2 shows, reasons for excluding samples include: contamination, coverage below the analysis threshold of 2000x in non-homopolymeric regions, and inconsistent haplogroup assignment and self-reporting of ethnicity resulting in the generation of European haplogroups.

Table 2. Summary information for inclusion and exclusion of samples for the African, Asian, and Latino population datasets.

	Sample Category	Number of Samples	Percentage of Samples
Included	**Satisfactory**	377	59.6%
Excluded	**Low coverage**	121	19.1%
	European	90	14.2%
	Contaminated	47	7.4%

The samples represented in table 2 resulted from 10 total sequencing runs; nine runs consisted of 72 samples each, and one run consisted of 64 samples. The author of this report performed the first eight of these sequencing runs. Low coverage was the reason resulting in the most excluded samples. Troubleshooting in attempt to remedy this issue included the adoption of normalization by quantification and dilution, rather than normalization by magnetic beads. Results of each previous run were considered in the following run to correct for consistency in library preparation methods and to target proper loading concentration to enrich for deep coverage.

Ultimately, samples which repeatedly provided inadequate coverage were excluded from further analysis. Samples excluded from analysis which presented more than one issue were counted once in this table.

Samples considered to be European were excluded from analysis because the target populations were African, Asian, and Latino. All European samples discussed in this report were result of previous sequencing and analysis (courtesy of Jennifer McElhoe, Ph.D.) and used strictly as comparison to the results generated for this report. Sample duplication was performed in an attempt to resolve low coverage or contamination which was not believed to have affected the extracted sample. Samples eliminated due to contamination included those in which the contamination affected the sample extract or source which could not be remedied by duplicate sequencing, and samples which were determined to be contaminated beyond the sequencing timeline for this report and could not be duplicated.

Table 3: Number of observations of no heteroplasmy, heteroplasmy, and minor allele frequency of 2-10% and greater than 10% for African, Asian, and Latino population groups.

	African	**Asian**	**Latino**
Total Samples	195	125	57
Samples without Heteroplasmy	130	83	45
Samples with Heteroplasmy	65	41	13
Samples with Heteroplasmy 2%-10%	50	31	9
Samples with Heteroplasmy >10%	15	10	4

Table 3 shows categorical results of the 377 samples deemed appropriate for reporting for this thesis exercise. Supplemental table 1 is a full list of analyzed samples including haplogroup, haplotype, and heteroplasmy information. Heteroplasmy information given in supplementary table

1 has the following format: T16093C 2.50, where the first letter (T) is the reference allele, the first number (16093) is the chromosome position, the second letter (C) is the variant allele, and the second number (2.50) is the minor allele frequency, or the rate at which the minor allele is present at that position. As DCMPS is useful to resolve heteroplasmy down to 2%, the samples with heteroplasmic observations were further characterized based on the minor allele proportion into two groups: minor allele frequency 2-10% and minor allele frequency >10%.

Figure 2: Haplogroup distribution for the African, Asian, and Latino population groups, as determined by HaploGrep2. In this figure n represents the number of samples belonging to each population.

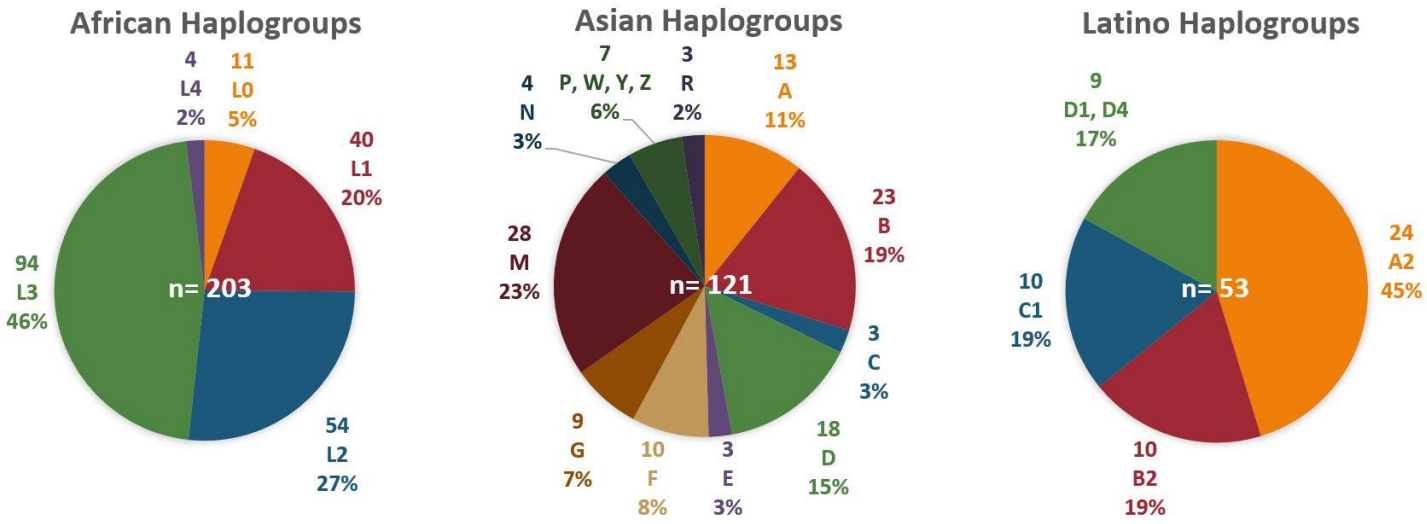


Figure 2 shows the sequence generated haplogroup results for the African, Asian, and Latino population groups. Of these samples, 177 self-reported as Black, African, or African-American; 91 self-reported as East Asian; 93 self-reported as Hispanic or Latino; and 16 self-reported as European. The inconsistency between reported ancestry and actual haplotype/haplogroup is the reliance on self-reporting ethnicity for sample collection.

Haplogrep2 haplogroup assignment considers 5437 haplogroups present in PhyloTree build 17 (van Oven, 2015). This dataset shows 202 haplogroups with 74 being shared by more than one sample and 128 being represented by one sample. Supplementary table 3 shows a full list of the haplogroups seen in this sample set with their population group assignment and number of observations. Haplogroup L3e2b+152 is shared by the greatest number of individuals at 11, and the related haplogroup L3e2b is shared by the second greatest number of individuals at 10. Three haplogroups were seen in seven individuals each, four were seen in six individuals, seven were seen in five individuals, seven were seen in four individuals, 18 were seen in three individuals, and 33 were seen in two individuals.

Table 4: Shared haplotypes including haplogroup designation and number of observations. Haplogroup information was generated using Haplogrep 2.

Haplogroup	Haplotype	Number of Observations
L3b	73G 263G 315.1C 523d 524d 16124C 16223T 16278T 16362C 16519C	4
L3e2b	73G 150T 195C 263G 315.1C 16172C 16183C 16189C 16223T 16320T 16519C	3
B2d	73G 263G 309.1C 315.1C 498d 499A 16183C 16189C 16217C 16519C	3
D1	73G 263G 315.1C 489C 16223T 16325C 16362C	3
L3e1e2	90A 97A 106d 107d 108d 109d 110d 111d 150T 189G 200G 263G 315.1C 16223T 16327T	3
L3f1b1	73G 189G 263G 315.1C 16183C 16189C 16209C 16223T 16292T 16295T 16311C 16519C	2
C1b2	73G 249d 290d 291d 315.1C 489C 493G 523d 524d 16223T 16298C 16325C 16327T 16519C	2
L3b1a+@16124	73G 263G 315.1C 523d 524d 16093C 16223T 16278T 16362C 16519C	2
L0a1b1	93G 95C 185A 189G 236C 247A 263G 315.1C 523d 524d 16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T	2

Haplogroup	Haplotype	Number of Observations
R9b1b	58C 73G 152C 263G 309.1C 315.1C 16124C 16148T 16304C 16309G 16327T 16390A 16519C	2
A2+(64)	64T 73G 146C 153G 235G 263G 309.1C 315.1C 523d 524d 16111T 16223T 16290T 16319A 16362C	2
L0a2a2	64T 93G 152C 189G 204C 207A 236C 247A 263G 315.1C 523d 524d 16148T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C	2
L0a1a2	64T 93G 185A 189G 200G 247A 263G 315.1C 514T 523d 524d 16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16362C	2
L2a1+143+@16309	73G 143A 146C 152C 195C 263G 315.1C 16129A 16223T 16278T 16294T 16390A	2
L2a1a2	73G 146C 152C 195C 263G 309.1C 315.1C 16223T 16278T 16286T 16294T 16309G 16390A 16519C	2
L2a1	73G 146C 152C 195C 263G 309.1C 315.1C 16223T 16278T 16294T 16309G 16390A 16519C	2
A2	73G 146C 153G 235G 263G 309.1C 315.1C 523d 524d 16111T 16223T 16290T 16319A 16362C	2
L3e2b+152	73G 150T 152C 195C 263G 309.1C 315.1C 16172C 16182C 16183C 16189C 16223T 16320T 16519C 16524G	2
L3e2b+152	73G 150T 152C 195C 263G 309.1C 315.1C 16172C 16183C 16189C 16223T 16320T 16519C	2
L3e3	73G 150T 195C 263G 315.1C 523d 524d 16093C 16223T 16265T 16519C	2
M7b1a1+(16192)	73G 150T 199C 263G 309.1C 315.1C 489C 16129A 16192T 16223T 16297C	2

Table 4 shows all haplotypes that were shared within this dataset. Of the 377 haplotypes, 350 different haplotypes (92.8%) were observed with 329 samples having unique (87.3%) haplotypes. The highest number of samples sharing one haplotype was 4. All haplotypes not included within table 4 were represented by only one sample and can be seen in supplementary table 1

Figure 3: Rate of heteroplasmic observations with minor allele frequencies of 2-10% and greater than 10%. In this figure, n represents the number of observations of heteroplasmy within the given category.

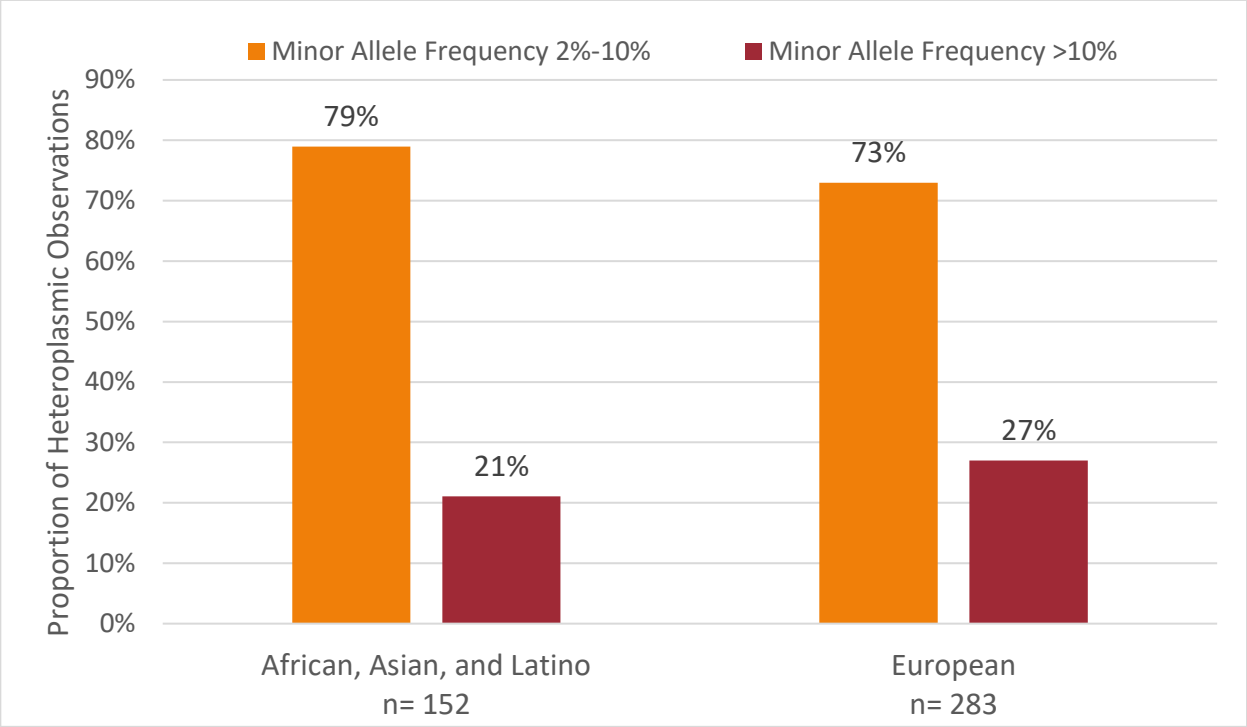
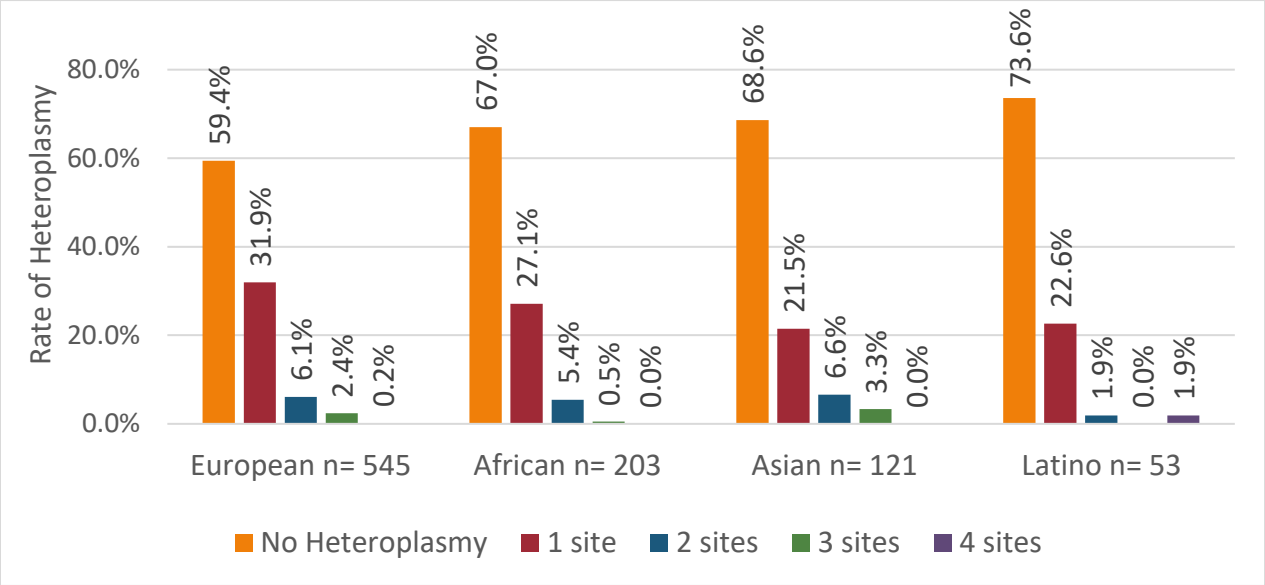


Figure 3 presents the proportion of heteroplasmic occurrences at based on minor allele frequency, which is the proportion of the minor allele at any given nucleotide position within an individual. When heteroplasmy occurs, there is a mixture of nucleotides at a single nucleotide position. The utility of DCMPS is the ability to resolve heteroplasmy below the accepted Sanger sequencing threshold of 10%. Therefore, the total number of heteroplasmic occurrences were grouped based on minor allele frequency of 2-10% and greater than 10%. The rate of heteroplasmy above 10% in the sample set of European, African, Asian, and Latino samples is 11.7% (108/922) with the rate of heteroplasmy above 2% being 47.2% (435/922).

Figure 4: Rates of heteroplasmy observed at ≥2% frequency based on the number of observations of heteroplasmy per individual for different population groups. In this figure, n represents the total number of samples per population group.



Rates of heteroplasmy within the European, African, Asian, and Latino population groups can be seen in figure 4. The rates were calculated by the dividing the number of samples with one, two, three, or four observations of heteroplasmy per individual by the total number of samples within that population. The European data set of 545 samples showed 324 (59.4%) samples with no heteroplasmy, 174 (31.9%) samples with one site of heteroplasmy, 33 (6.1%) samples with two sites of heteroplasmy, 13 (2.4%) samples with three sites of heteroplasmy and one (0.2%) sample with four sites of heteroplasmy. The African dataset of 203 samples showed 136 samples with no heteroplasmy (67.0%), 55 samples with one site of heteroplasmy (27.1%), 11 samples with two sites of heteroplasmy (5.4%), one sample with three sites of heteroplasmy (0.5%), and no samples with four sites of heteroplasmy. The Asian dataset with 121 samples showed 83 samples with no heteroplasmy (68.6%), 26 samples with one site of heteroplasmy (21.5%), 8 samples with two sites of heteroplasmy (6.6%), four samples with three sites of heteroplasmy (3.3%), and no samples

with four sites of heteroplasmy. The Latino dataset with 53 samples showed 39 samples with no heteroplasmy (73.6%), 12 samples with one site of heteroplasmy (22.6%), one sample with two sites of heteroplasmy (1.9%), no samples with three sites of heteroplasmy and one sample with four sites of heteroplasmy (1.9%).

Table 5: P-value (α=0.05) results of two sample z-test for proportions (no Yate's correction) comparing rates of the presence of heteroplasmy in African, Asian, Latino, and European populations.

	African	**Asian**	**Latino**
Asian	0.766	X	X
Latino	0.358	0.508	X
European	0.059	0.062	*0.044*

Table 5 shows p-values results of two-proportions z-tests (Wilson, 1927) for equality of proportions without continuity correction that were applied to evaluate the statistical significance of the difference in rates of heteroplasmy when comparing across two population groups. The alpha value for the z-test was set at 0.05. Tests were performed on RStudio, version 1.1.383 (RStudio Team, 2016). The comparison between European and Latino populations produced the only significant p-value at 0.044.

Figure 5: Expected rate of heteroplasmy based on a 95% confidence interval around the rate of heteroplasmy for European, African, Asian, and Latino populations in this study.

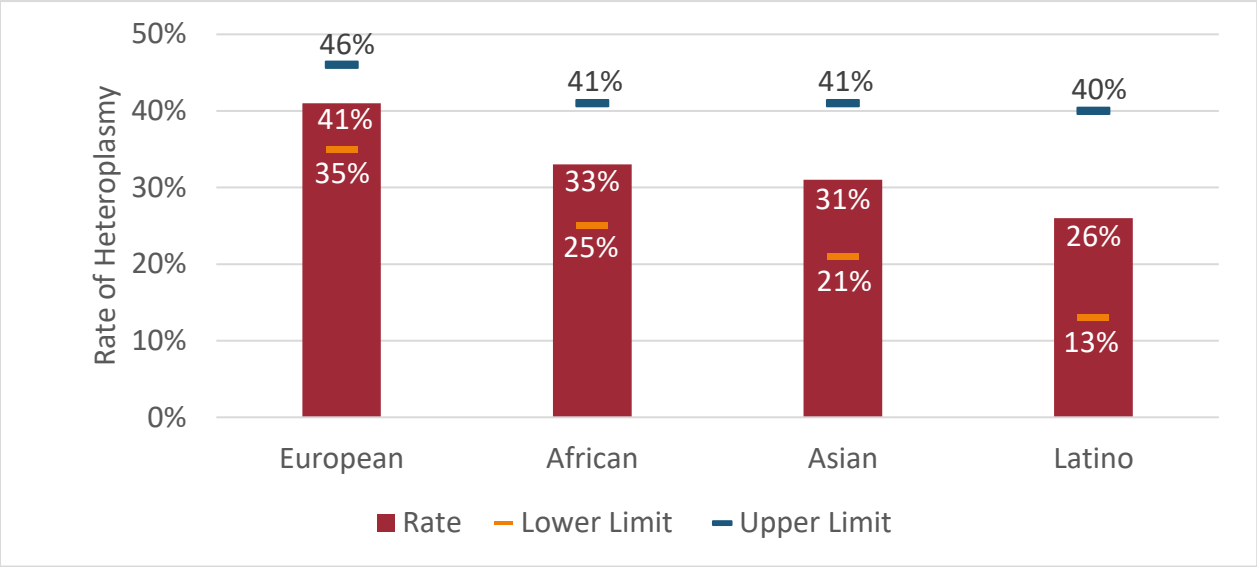


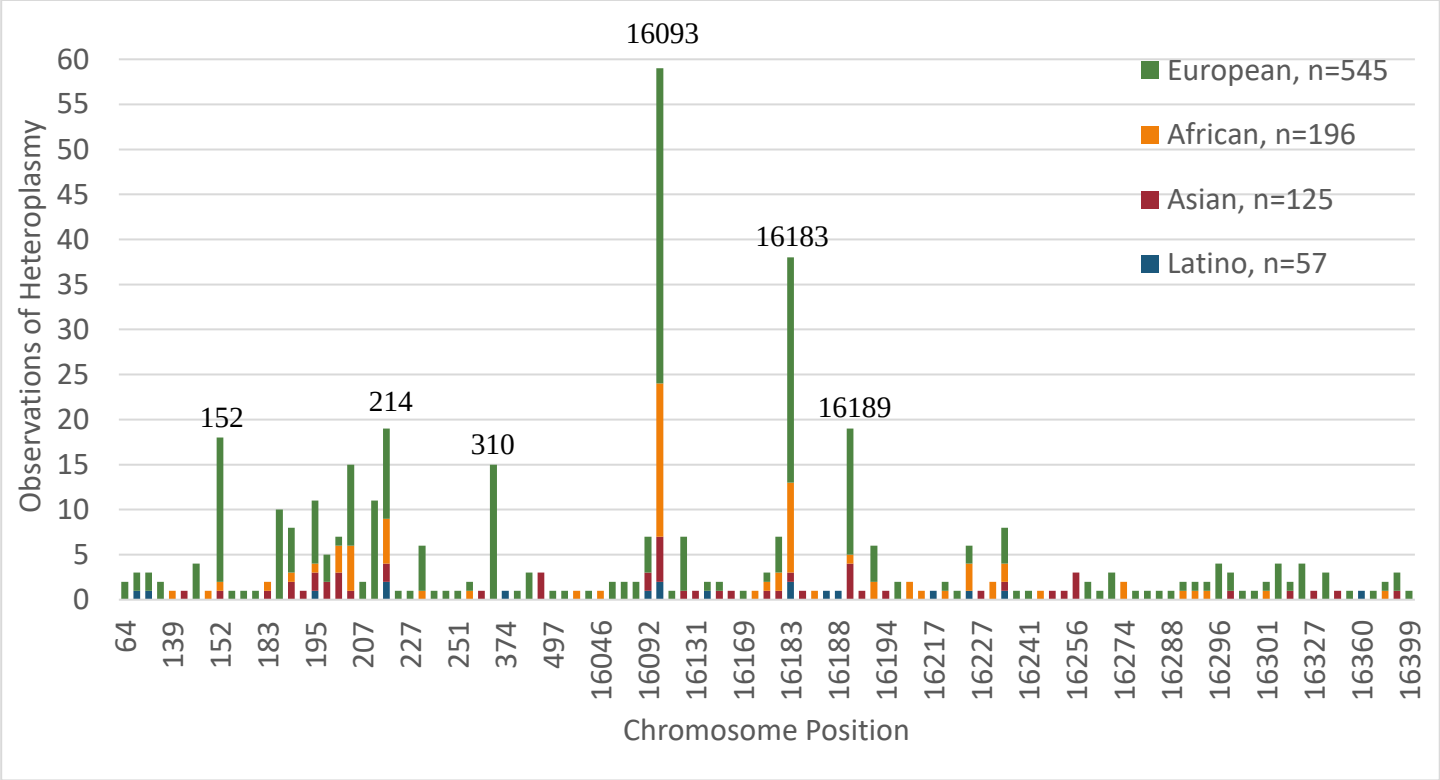
Figure 5 shows the calculated 95% confidence interval around the rate of heteroplasmy. The rate presented was the rate of heteroplasmy based on the European (n=545), African (n=203), Asian (n=121), and Latino (n=53) populations in this study. The upper and lower limits were calculated using:

$$\text{Upper limit} = \frac{100}{n} * (d + 1.96 * \sqrt{d})$$

$$\text{Lower limit} = \frac{100}{n} * (d - 1.96 * \sqrt{d})$$

where *n*= denominator of the rate (i.e. sample size) and *d*= number of events upon which the rate is based (ie. number of samples with observations of heteroplasmy).The 100 is based on our rate calculations where the rate equals the number of individuals with heteroplasmy divided by the total number of observations multiplied by 100 (i.e. European population rate = 221/545*100 = 40.6%). Therefore the rate can also be stated as 40.6 observations of heteroplasmy per 100 individuals.

Figure 6: Observations of heteroplasmy across the CR by population group.



The observations of heteroplasmy across the mtDNA CR for all combined data can be seen in figure 6, with supplementary table 2 outlining the data. The x-axis represents the mtDNA chromosome position starting with nucleotide position 1-576 followed by positions 16024-16569.

Figure 6 is not normalized to population and represents only the positions at which heteroplasmy occurred across the control region. Rates of heteroplasmy are calculated by dividing the number of observations per position by the total number of individuals within the given population. Combined, the four populations give a total of 435 heteroplasmic observations across 114 sites. The European population shows 283 observations at 81 sites, African shows 78 observations at 35 sites, Asian shows 56 observations at 39 sites, and Latino shows 18 observations

at 15 sites. Out of total observations of heteroplasmy, 56 sites appeared once in the dataset, 23 appeared twice, 10 appeared 3 times, 4 appeared four times, and 21 sites appeared 5 or more times.

Figure 7: Rate of heteroplasmy for the most common positions by population group.

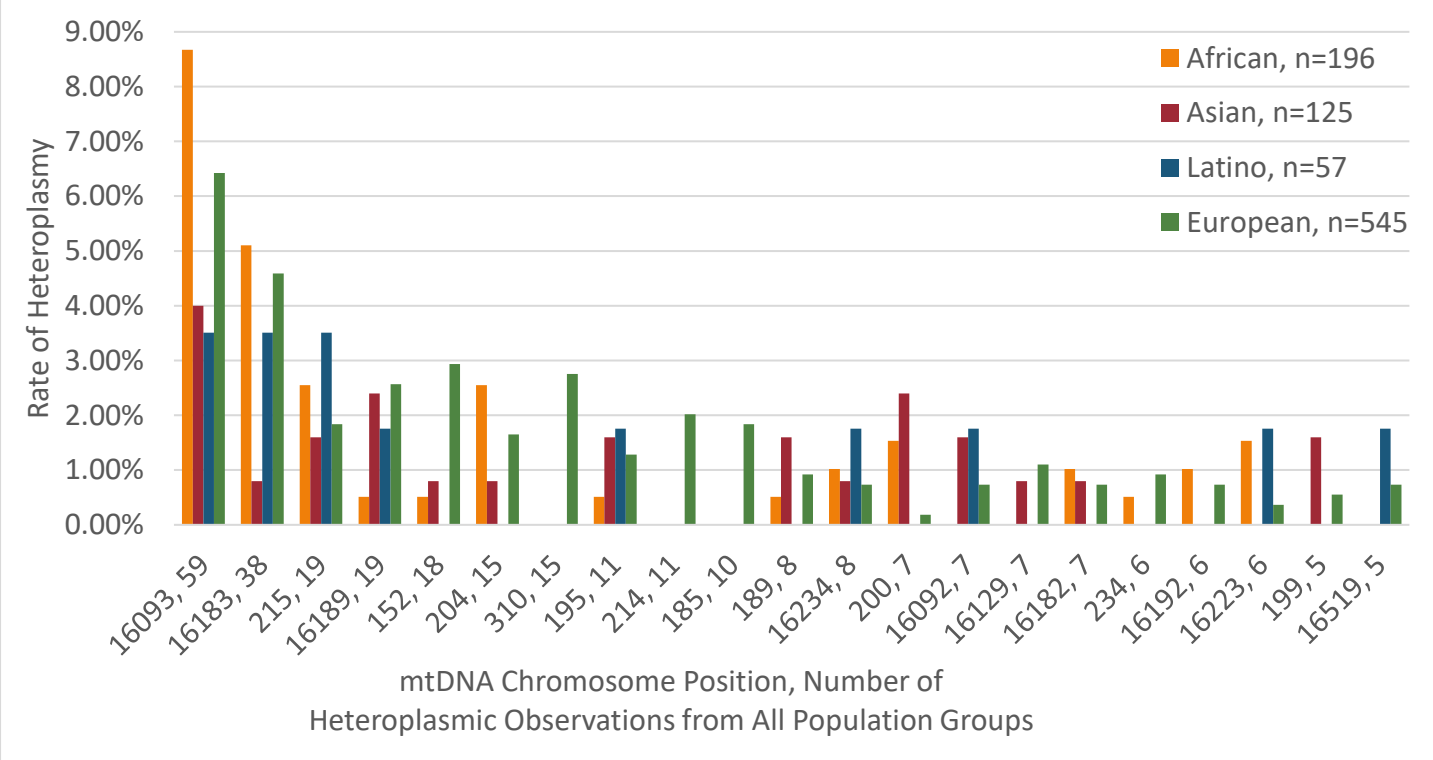


Figure 7 includes the rate of heteroplasmy among the four populations at all CR positions which exhibited 5 or more observations of heteroplasmy. The x-axis shows the mtDNA chromosome position number followed by the number of observations for that position. The rate of heteroplasmy was determined by dividing the number of observations at a given position by the total number of samples within that population.

Discussion

The sequencing method in this report has proven to be robust in previous studies (Gallimore et al, 2018; Rathbun et al, 2107; Holland et al, 2018) and in the sequencing of the European dataset, however low coverage has diminished the proportion of useable data for this report. The cause of inadequate coverage is not fully known for these samples. From the third run forward, manual library normalization was adopted to reduce user error in bead normalization and to ensure greater consistency within the bench work. This change did not appear to have a quantifiable effect. Duplicated samples often showed low coverage among the same CR positions. The third run had the greatest number of samples with low coverage issues. This run consisted largely of samples stored frozen from collection in 2014. It is unknown if the storage of these samples caused these issues, and further testing of stored samples should be performed to investigate this result.

It should be stated that while a small number of samples showed low coverage across the entire control region, the most common pattern of depleted coverage occurred in homopolymeric and repeat sequence stretches; including nps 16184-16193, 303-315, and 514-524. Length polymorphisms within these regions are common and therefore have low discrimination potential, so they are often not considered in mtDNA testing (SWGDAM, 2013). Low resolution in these areas is not wholly detrimental to the DCMPS approach. Overall, completion of proposed 750 non-European mtDNA sequences should assist in the troubleshooting of this drawback.

Contamination was the other factor reducing results which could affect forensic applications of this method. The low percentage of human contamination was a result of careful benchwork, separate pre-PCR and post-PCR laboratories, and strict anti-contamination cleaning

procedures. Contamination was kept below 4 samples per run for the first nine runs. Often, contamination deemed low-level (showing a low number of minor variants that could be contributed to contamination) could be remedied by re-sequencing, which indicated run-specific contamination on the part of the researcher. The tenth run saw contamination elevated to 47 samples, many of which showed high level contamination characterized by roughly 10 or more minor variants contributable to contamination. These samples were resequenced after the course of this report and no contamination appeared, leading to the conclusion that run-specific preparation was the cause for this elevated contamination.

Haplotype data was concordant with expectations. Relatively few haplotypes were shared among samples sequenced, as the majority of haplotypes (94%, 329/350) were unique. More haplogroup designations were shared among samples, which is expected as haplogroups are related groupings of haplotypes. Haplogroup distributions were not unexpected for sequencing self-reported samples. The African samples were comprised of 81% who self-reported as Black, African, or African American. The Asian samples were comprised of 71% who self-reported as East Asian. The Latino samples were comprised of 91% who self-reported as Hispanic or Latino. The haplogroup designations matched the sorting expectations listed within the methods. All samples which did not match those listed in the methods were within the haplogroups designated to be European in the previous study. Loss of self-reported Hispanic or Latino samples occurred when the result was often African or European, based on the admixed nature of Latino populations. Also, some self-reported Hispanic or Latino samples gave Asian haplogroup results of types associated with basal nodes, rather than the Native American branches of the sequences. The African and Asian populations both showed 14% of samples who self-reported as Hispanic or Latino. Haplogroup sorting is crucial to this study because haplogroups consists of naturally

clustered haplotypes which arise from similar lineages. Samples sorted into macrohaplogroup clades provide clusters of samples with similar mtDNA sequences and can be used to analyze if heteroplasmy may be linked to sequence-dependent groupings.

Overall, this report illustrated that all population groups tested have moderate rates of heteroplasmy ranging from 26% in the Latino group to 41% in the European group. According to two proportion z-tests, the only two-way population group comparison showing significant difference in rate of heteroplasmy is Latino vs. European. However, this slight significance must be considered carefully due to the small Latino sample size. If 100 random individuals of the populations tested within this report were sequenced using the presented method, then the number of individuals expected to present heteroplasmy greater than 2% can be seen in figure 5. When heteroplasmy does occur, it is roughly 3 times more likely to be between 2-10% than above 10%, as shown by figure 3. This leads to the conclusion that DCMPS should be adopted in the interest of improving the resolution of low-level heteroplasmic variants.

The overall rate of heteroplasmy using DCMPS, for all populations combined is 36.8%, which is roughly 6 fold higher than previously reported rates of heteroplasmy by Sanger (6%) sequencing (Irwin et al, 2009). The European dataset showed the highest rate of heteroplasmy at 40.6%. The African rate of heteroplasmy was 33%, Asian was 31.4%, and Latino was 26.4%. It should be noted that the rate of heteroplasmy decreased with decreasing sample size. Perhaps, with the addition of more samples, the rate of heteroplasmy would increase and the rate of heteroplasmy may become more similar across populations.

Overall, all population groups show the highest rate of heteroplasmy for one site per individual, with additional sites becoming increasingly rare. This is evidence for the purifying process in mitochondria which filters out mutations (Greaves et al, 2014). Only two samples in the

combined 922 samples showed four sites of heteroplasmy, making this rate 0.22%. The rarity of multiple sites of heteroplasmy also allows for ease in contamination identification. Any sample with heteroplasmic sites requires careful scrutiny to ensure that all sites are reliable. More than four sites of heteroplasmy within one sample can be considered a warning that contamination may have occurred, and a sign that deep analysis of MPS read data is required to validate true heteroplasmic positions.

Position 16093 appears to be a hot spot for heteroplasmy in the mtDNA CR for all four populations, which agrees with previously reported data (Irwin et al, 2009). In each population it is the position with the most observations of heteroplasmy. T16093C is a fairly common SNP seen in all macro population groups (Parson and Dür, 2007). Perhaps the common polymorphic state of this position drives the heteroplasmy. The T16093C polymorphism occurred in 21 of the 377 African, Asian, and Latino haplogroup-based populations. PHP at this position occurs both as the non-rCRS reference minor allele C (4 observations, 1% African, Asian, Latino combined population rate), and as the rCRS reference minor allele T (20 observations, 5.3% African, Asian, Latino combined population rate). This site is not in a homopolymeric region, so the high rate of heteroplasmy observed here is possibly a visualization of mtDNA sequences transitioning between a GC or AT base pair at this position. As the reference sequence is of European origin (Anderson et al., 1981), and T16093C occurs in all macropoulations, no conclusions can be made that heteroplasmy at 16093 is linked to any one population.

Other apparent hot spots include 16183 and 16189, both with considerable numbers of occurrences in multiple populations. The rCRS (Andrews et al., 1999) shows a homopolymeric C stretch at nps 16184-16193, interrupted by a single T at 16189. T16189C is a common SNP in all macro populations with slightly higher frequency in Africans (Parson and Dür, 2007) which then

causes a 10 C stretch that often presents with length heteroplasmy. A16183C is considered point heteroplasmy by the forensic community (Bandelt and Parson, 2008) however, it can also be hypothesized to be an A deletion alongside the insertion of a C. Ultimately, there is no way to empirically prove if the point transversion or the indel occurred in this polymorphism, and forensic genetic alignments favor Occam's Razor in reporting the option that requires the least number of assumptions. Homopolymeric regions and repeated sequences, like 161484-16193 and 303-310 should consider elevated point heteroplasmy as a possible result of length heteroplasmic changes.

Position 215 may also be a hot spot for heteroplasmy, with figure 4 showing a possible warm zone for heteroplasmy occurring between nps 152-215. Besides the hot spots discussed above, this zone shows the most occurrences of heteroplasmy in a more concentrated area of the mtDNA CR. Interestingly, This area is linked to the origin of heavy-strand synthesis for mtDNA replication (Fish et al, 2004; Nicholls and Minczuk, 2014). Origin positions 151, 168, and 191 (Fish et al, 2004; Nicholls and Minczuk, 2014) do not show heteroplasmy within this report, but the apparent higher frequency of heteroplasmic positions will be interesting to investigate possible significance. Perhaps frequent variations in this small area may be a contributing factor to the multiple origin site replicative technique of mtDNA. Possibly, as new variations arise, so do new lengths of 7S DNA involved in D-loop structure and function.

Positions where one observance of heteroplasmy occurred appear spread evenly across the mtDNA CR, causing non-hot spots to have relatively equal rates of observations. The single observations of heteroplasmy will be difficult to observe any connection to sequence variants, so they should be considered random until further testing is able to suggest otherwise.

Nucleotide positions nps 310, 214, and 185 interestingly appear often as heteroplasmy in the European population but do not appear in the African, Asian, and Latino populations. T310C

is a SNP which causes an extended C stretch from 303-315 and occurs often in European, Asian, and Latino populations but less frequently in the African population (Parson and Dür, 2007). It is possible that heteroplasmy at 310 was not seen in the current dataset due to sampling limitations and the issues experienced with low coverage. Replication difficulties associated with γ-polymerase and PCR slippage in extended homoplymeric regions can make these regions rife with in vivo and in vitro variation and therefore difficult to align and report. Samples which contained low coverage only in the 310 area were still included in this report in the interest of increasing results and due to this area often being excluded from reporting protocols based on low discrimination potential.

Positions 214 and 185 have less simple possible explanations. The non-European dataset is 69% of the size of the European dataset, so possibly these positions may show heteroplasmy in future sequencing results. However, these positions are not in a homopolymeric stretch. In our dataset, A214G appears in Asian and Latino populations, and according to EMPOP, it is fairly common among Americans but seen in all other metapopulations. G185A appears in African samples in this study and has the highest frequency in African and European populations on EMPOP. These SNPs may be driving heteroplasmy in the European populations because overall heteroplasmy is showing to occur more in this population than in those newly sequenced. Especially with 185, the position may be in a transitional state. African sequences are more basal than European, and European sequences are more basal than Asian and Latino, so possibly European individuals are presenting heteroplasmy at 185 as the position switches from an ancestral A, which occurred in African lineages and happened to not be a part of the rCRS and is mutating to G in the younger lineages.

Overall, the results of this study show that heteroplasmy occurs in 36% of the general population, with variance across European, African, Asian, and Latino population groups that may be explained by sample size variation. Also, when heteroplasmy does occur within an individual, it is most likely that the individual will have a single site of heteroplasmy and that site will have a frequency below 10% regardless of population group. Nucleotide position 16093 shows the highest rate of heteroplasmy within the CR, and there may be increased heteroplasmy in homopolymeric regions as well as in the area surrounding the heavy strand origin of replication. MPS sequencing of the mtDNA CR using this protocol was shown to be successful and analysis by GeneMarker® HTS and HaploGrep2 are fairly straightforward. Results such as those presented here are achievable in a Forensic Science setting.

With a full dataset, the next steps will be to assess possible linkage between nucleotide sequence and heteroplasmy. It is believed that if this occurs, it may be along haplogroup lines, as these groups are based upon related ancestral lineages. This linkage assessment is the reason why this data is presented with haplogroup based population groups, rather than grouping samples based on their self-reported ancestry. In the forensic field, an evidentiary sample will not have self-reported data. Sequencing will need to occur before a possible population group can be assigned. The goal of this research is to understand the occurrences and rates of heteroplasmy among populations as though they are forensic samples. The case of Tsar Nicholas Romonov has proven that the occurrence of heteroplasmy can be a statistical strength in the case of a match. We now need to present reliable data to better calculate the possible weight of the newly resolvable levels of heteroplasmy that we have observed within this report.

Appendix

Supplemental Data

Supplementary Table 1: Sample list with haplogroup, haplotype, and heteroplasmy.

Sample ID	Haplogroup	Haplotype	Heteroplasmy
28	M7c1c2	73G 146A 199C 204C 263G 309.1C 315.1C 489C 523d 524d 16519C	
29	F1a3+16311	73G 152C 249d 263G 315.1C 523d 524d 16129A 16172C 16304C 16311C 16497G 16519C	C16488T 2.27
30	C1c4	73G 214G 249d 263G 290d 291d 309.1CC 315.1C 489C 16223T 16298C 16325C 16327T 16519C	
38	D4a8	73G 146C 152C 263G 309.1CC 315.1C 489C 523d 524d 16129A 16223T 16270T 16362C 16519C	
45	D1	73G 196C 263G 315.1C 489C 16223T 16325C 16362C	
57	G2a1c	73G 263G 309.1C 315.1C 489C 16093C 16183C 16189C 16194G 16195C 16223T 16227G 16278T 16362C 16519C	A16194C 12.36 T16093T 3.06
93	A2+(64)	64T 73G 146C 153G 235G 263G 309.1CC 315.1C 523d 524d 16111T 16223T 16290T 16319A 16519C	
277	M7c1	73G 146C 199C 263G 315.1C 489C 523d 524d 16223T 16295T 16519C	
288	D5b1c1	73G 150T 152C 185A 263G 315.1C 456T 489C 523d 524d 16092C 16148T 16183C 16189C 16362C 16519C	
311	B5b2c	73G 103A 131C 146C 263G 309.1C 315.1C 481T 523d 524d 16111T 16140C 16183C 16189C 16234T 16243C 16463G 16519C	
317	M9	73G 263G 309.1C 315.1C 489C 16223T 16362C 16519C	
408	L4b1a	73G 150T 199C 204C 263G 309.1C 315.1C 513A 523d 524d 16182C 16183C 16189C 16223T 16239T 16311C 16320T 16362C 16519C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
458	M7c1a3a	61T 62A 73G 146C 199C 228A 257G 263G 315.1C 489C 513A 523d 524d 16223T 16319A 16519C	
472	L1b	73G 152C 182T 185T 195C 247A 263G 315.1C 357G 523d 524d 16126C 16129A 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C	
493	D4e1	73G 263G 315.1C 328G 489C 16092C 16223T 16362C 16519C	T16092T 5.34
625	B4a1a	73G 146C 263G 309.1CC 315.1C 523d 524d 16182C 16183C 16189C 16217C 16261T 16304C 16519C	
740	L3e2b	73G 150T 195C 263G 315.1C 316A 16172C 16189C 16223T 16320T 16519C	
741	L2c3	73G 89C 93G 146C 150T 182T 195C 198T 263G 309.1C 315.1C 325T 513A 523d 524d 16223T 16278T 16390A	
742	L2b	73G 146C 150T 152C 182T 195C 198T 204C 207A 263G 315.1C 16114A 16129A 16212G 16213A 16223T 16278T 16390A	
743	L2b	73G 146C 150T 152C 182T 183G 195C 198T 204C 263G 309.1C 315.1C 16114A 16129A 16213A 16223T 16274A 16278T 16390A	
744	L1c2a1	73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 315.1C 316A 523d 524d 16129A 16187T 16189C 16213A 16223T 16265C 16278T 16286G 16294T 16311C 16360T 16519C 16527T	
745	L1b	73G 152C 182T 185T 195C 247A 263G 315.1C 357G 523d 524d 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16311C 16519C	
747	L1b1a+189	73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523d 524d 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C	
748	L3h1b1a	73G 189C 195C 263G 309.1C 315.1C 523d 524d 16179T 16215G 16223T 16256A 16284G 16311C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
749	L3f1b1	73G 189G 195C 200G 263G 315.1C 16209C 16223T 16292T 16295T 16311C 16519C	T204C 2.53
752	L2b2a	73G 146C 150T 152C 182T 195C 198T 204C 263G 315.1C 16114A 16129A 16213A 16223T 16278T 16354T 16390A	
755	A2ai	64T 73G 146C 153G 235G 263G 309.1C 315.1C 523d 524d 16213A 16223T 16290T 16319A 16362C	
757	L3d1b3a	73G 146C 152C 263G 315.1C 523d 524d 16124C 16223T 16256T	
758	L3e2b	73G 150T 195C 263G 315.1C 16172C 16189C 16223T 16320T 16519C	
760	L2a1	73G 146C 152C 195C 263G 309.1C 315.1C 16189C 16223T 16278T 16294T 16309G 16390A	
761	D4g2	73G 263G 298T 309.1CC 315.1C 489C 16223T 16362C 16519C	
764	L3e2a1b1	73G 150T 195C 198T 263G 315.1C 499A 16140C 16223T 16320T 16399G 16519C	
765	L3e1a	73G 150T 152C 189G 200G 207A 263G 309.1C 315.1C 16111T 16185T 16223T 16264T 16319A 16327T	
767	L3e2b+152	73G 150T 152C 195C 263G 309.1C 315.1C 16172C 16183C 16189C 16223T 16320T 16519C	
769	L1c2a1a	73G 146C 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 315.1C 316A 16071T 16129A 16145A 16187T 16189C 16213A 16223T 16234T 16265C 16278T 16286A 16294T 16311C 16360T 16527T	C16234C 2.06
770	L2a1	73G 146C 152C 195C 263G 315.1C 16093C 16189C 16223T 16278T 16294T 16309G 16390A 16519C	A16183C 3.06 T16093T 2.27
772	L3d1b2	73G 150T 152C 195C 263G 315.1C 523d 524d 16124C 16223T 16292T	
774	B4+16261	73G 263G 309.1C 315.1C 16182C 16183C 16189C 16217C 16227G 16261T 16299G 16355T 16390A 16519C	A16227A 2.6 C16148T 6

Sample ID	Haplogroup	Haplotype	Heteroplasmy
776	Z4	73G 151T 152C 249d 263G 309.1C 315.1C 489C 16086C 16129A 16185T 16223T 16260T 16298C	
778	N9a1'3	73G 150T 263G 309.1C 315.1C 16129A 16223T 16257A 16261T	
780	L1c3b'c	73G 151T 152C 182T 186A 189C 247A 263G 265C 315.1C 316A 523d 524d 16129A 16187T 16189C 16223T 16278T 16293G 16294T 16311C 16360T 16519C 16527T	
781	L3d5	73G 152C 199C 263G 315.1C 523d 524d 16124C 16223T 16362C 16519C	
782	D4g2	73G 263G 298T 309.1C 315.1C 489C 16177G 16223T 16362C 16526A	
784	L2a1+143+@16309	73G 143A 146C 152C 195C 263G 315.1C 16223T 16278T 16294T 16390A 16519C	
787	B2o	73G 159C 263G 315.1C 499A 16092C 16182C 16183C 16189C 16193d 16217C 16519C	G16145A 5.98
788	L3e3	73G 150T 195C 263G 315.1C 523d 524d 16223T 16265T 16519C	
789	F1a1c	73G 249d 263G 309.1C 315.1C 523d 524d 548T 16129A 16172C 16184T 16304C 16519C	
790	L1b	73G 152C 182T 185C 195C 247A 263G 315.1C 357G 523d 524d 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C	A263A 21.56
791	L2b	73G 146C 150T 152C 182T 195C 198T 204C 207A 263G 315.1C 16114A 16129A 16213A 16223T 16278T 16390A	
792	L0a1a2	64T 93G 185A 189G 200G 247A 263G 315.1C 514T 523d 524d 16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16362C	
793	L3e2b	73G 150T 195C 263G 315.1C 16172C 16183C 16189C 16223T 16320T 16519C	
795	F1a1d	73G 249d 263G 309.1C 315.1C 523d 524d 16129A 16162G 16172C 16304C 16399G 16519C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
796	M9a1a1c	73G 153G 263G 309.1C 315.1C 482C 489C 16223T 16234T 16291T 16316G 16362C	T16131C 2.36 A215G 7.47 C16234C 2.05
797	D4b2b	73G 194T 263G 309.1C 315.1C 489C 523d 524d 16223T 16362C 16519C	T16189C 10.86
798	L0d1'2	73G 146C 152C 195C 247A 315.1C 316A 498d 16129A 16187T 16189C 16223T 16230G 16243C 16311C 16519C	T16093C 9.74
801	N9a1	73G 150T 263G 309.1C 315.1C 16111T 16129A 16223T 16257A 16261T	
802	C1b4	73G 143A 196d 249d 263G 290d 291d 309.1C 315.1C 489C 493G 523d 524d 16086C 16183C 16189C 16223T 16278T 16298C 16325C 16327T	
803	G2c	73G 152C 195C 263G 309.1C 315.1C 489C 16223T 16362C	
805	L3e2b+152	73G 150T 152C 195C 263G 309.1C 315.1C 16172C 16182C 16183C 16189C 16223T 16320T 16519C 16524G	
806	L3e2b+152	73G 150T 152C 195C 263G 309.1C 315.1C 16172C 16182C 16183C 16189C 16223T 16320T 16519C 16524G	
809	D4a	73G 263G 309.1CC 315.1C 489C 523d 524d 16129A 16223T 16362C 16519C	
811	L1c3b	73G 93G 151T 152C 182T 186A 189C 247A 263G 315.1C 316A 523d 524d 16129A 16163G 16189C 16223T 16278T 16293G 16294T 16311C 16360T 16519C	T195C 3.83 A16183C 2.59
812	M71a1a	73G 150T 204C 263G 309.1C 315.1C 489C 16223T 16269G 16271C	
813	L2a1	73G 146C 152C 195C 263G 309.1C 315.1C 16223T 16278T 16294T 16309G 16390A 16519C	
814	L3b1a+@16124	73G 263G 315.1C 523d 524d 16223T 16278T 16311C 16362C 16519C	T16209C 3.4
816	L1c1d	73G 151T 152C 182T 186A 189C 195C 247A 263G 297G 315.1C 316A 480C 523d 524d 16038G 16086C 16129A 16187T 16189C 16192T 16278T 16293G 16294T 16360T 16519C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
817	C1b2	73G 249d 290d 291d 315.1C 489C 493G 523d 524d 16223T 16298C 16325C 16327T 16519C	
818	A2	73G 146C 153G 235G 263G 309.1C 315.1C 523d 524d 16111T 16223T 16290T 16319A 16362C	
820	B4b	73G 263G 309.1CC 315.1C 499A 16129A 16183C 16189C 16217C 16362C 16519C	
822	L3e1e2	90A 97A 106d 107d 108d 109d 110d 111d 150T 189G 200G 263G 315.1C 16223T 16327T	
823	L3e1e2	90A 97A 106d 107d 108d 109d 110d 111d 150T 189G 200G 263G 315.1C 16223T 16327T	A200A 2.4
824	M7c1c2	73G 146A 199C 204C 263G 315.1C 489C 523d 524d 16519C	
826	F1a	73G 249d 263G 309.1C 315.1C 523d 524d 16129A 16172C 16304C 16519C	
827	D4a3a2	73G 152C 237G 263G 309.1C 315.1C 489C 16093C 16129A 16223T 16249C 16292T 16362C	T16093T 5.6
828	L3b1a+@16124	73G 263G 315.1C 523d 524d 16093C 16223T 16278T 16362C 16519C	A215G 19.82
830	C1b2	73G 249d 290d 291d 315.1C 489C 493G 523d 524d 16223T 16298C 16325C 16327T 16519C	
831	L3e1b2	73G 150T 185A 189G 263G 309.1C 315.1C 16223T 16325d 16327T	
832	L1c2	73G 150T 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 315.1C 316A 16129A 16166G 16187T 16189C 16223T 16230G 16265C 16278T 16284G 16286A 16294T 16311C 16519C 16527T	C151C 11.22
835	L3e1a	73G 150T 189G 200G 263G 315.1C 16185T 16223T 16327T 16519C	A200A 2.48
837	L3e2b	73G 150T 195C 263G 309.1C 315.1C 16172C 16183C 16189C 16223T 16320T 16519C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
839	B2h	73G 263G 315.1C 499A 16183C 16189C 16217C 16468C 16519C	A16183A 4.62 C16187T 3.86 T16217T 3.05 C16223T 2.01
841	L1c3a	73G 151T 152C 182T 186A 189C 247A 263G 315.1C 316A 523d 524d 16129A 16145A 16189C 16215G 16223T 16278T 16294T 16311C 16360T 16519C	
850	L1c1d	73G 151T 152C 182T 186A 189C 195C 198T 204C 247A 263G 297G 309.1C 315.1C 316A 523d 524d 16038G 16086C 16129A 16187T 16189C 16223T 16278T 16293G 16294T 16360T 16519C	
854	L3e1	73G 150T 189G 263G 315.1C 16179T 16223T 16327T 16519C	C16179C 2.53 C16327C 2.11
856	L0d3b	73G 146C 150T 195C 247A 315.1C 316A 16187T 16189C 16223T 16230G 16243C 16274A 16278T 16290T 16311C 16519C	
857	B4	73G 263G 309.1C 315.1C 498d 499A 16183C 16189C 16217C 16519C	
858	G3	73G 152C 263G 315.1C 489C 16223T 16274A 16362C	T195C 28.5
861	A14	73G 151T 152C 200G 235G 248G 263G 315.1C 523d 524d 16223T 16290T 16319A 16362C	A200A 4.43
862	L2a1+143	73G 143A 146C 152C 195C 263G 315.1C 523d 524d 16223T 16278T 16294T 16309G 16390A	
863	B5b2+@204	73G 103A 131C 146C 263G 309.1CC 315.1C 316A 523d 524d 16111T 16140C 16183C 16189C 16234T 16243C 16463G 16519C	
865	A14	73G 151T 152C 200G 235G 263G 309.1C 315.1C 523d 524d 16223T 16256T 16290T 16319A 16362C	
868	A2h	54C 64T 71d 73G 146C 153G 194T 235G 263G 315.1C 523d 524d 16111T 16175G 16185T 16290T 16319A 16362C 16526A	
869	L2b	73G 146C 150T 152C 182T 195C 198T 204C 263G 315.1C 513A 16114A 16129A 16213A 16223T 16265G 16278T 16311C 16368C 16390A	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
870	M7b1a1+(16192)	73G 150T 199C 263G 309.1C 315.1C 489C 16129A 16192T 16223T 16297C	G16129G 2.28 T199T 2.38 T16297T 2.55
871	L2b1a	73G 150T 152C 182T 195C 198T 204C 263G 292A 315.1C 418T 523d 524d 573.1C 16114A 16213A 16223T 16278T 16355T 16362C 16390A	
872	L3b	73G 263G 315.1C 523d 524d 16124C 16223T 16278T 16344T 16519C	
873	F4a1b	73G 146C 249d 263G 315.1C 317A 16126C 16189C 16207G 16304C 16362C 16399G	
874	L2a1a2	73G 146C 152C 195C 263G 309.1C 315.1C 16223T 16278T 16286T 16294T 16309G 16390A 16519C	T16093C 48.84
875	P4b	73G 210G 263G 309.1C 315.1C 523d 524d 16066G 16140C 16183C 16189C 16266G 16274A 16291T 16519C	
876	L0a1b1	93G 95C 185A 189G 236C 247A 263G 315.1C 523d 524d 16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T	A16230A 10.56 C16223C 7.36
877	L0a2a2	64T 93G 152C 189G 204C 207A 236C 247A 263G 315.1C 523d 524d 16148T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C	
878	L3e2b	73G 150T 195C 263G 315.1C 16172C 16189C 16223T 16320T 16519C	A16183C 2.84
879	L3e1f1a	73G 150T 189G 263G 315.1C 16183C 16189C 16223T 16260T 16327T	
880	L3e3	73G 150T 195C 263G 315.1C 523d 524d 16093C 16223T 16265T 16519C	T16209C 3.69 T16093T 4.88
881	L3e3	73G 150T 195C 263G 315.1C 523d 524d 16093C 16223T 16265T 16519C	T16093T 9.92
882	B2c2	63C 64T 73G 143A 146C 189G 215G 263G 309.1CC 315.1C 455.1T 499A 16183C 16189C 16214T 16217C 16519C	G66A 2.02
884	L2c2	73G 93G 146C 150T 152C 182T 195C 198T 263G 309.1C 315.1C 325T 523d	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
		524d 16192T 16223T 16264T 16278T 16390A	
885	D1	73G 263G 315.1C 489C 16223T 16325C 16362C	
886	A2+(64)+16129	64T 73G 125C 127C 146C 153G 235G 263G 309.1C 315.1C 499A 523d 524d 16111T 16129A 16223T 16290T 16319A 16354T 16362C 16519C	T16093C 22.5
887	A2+(64)+@153	64T 73G 146C 150T 235G 263G 315.1C 523d 524d 16111T 16223T 16290T 16319A 16362C 16519C	A374G 4.08 C16234T 25.73
888	L3d1b2	73G 150T 152C 263G 315.1C 523d 524d 16124C 16223T 16399G	
889	L4b2b1	73G 146C 152C 195C 244G 263G 309.1C 315.1C 340T 523d 524d 16051G 16114T 16189C 16192T 16223T 16293T 16311C 16316G 16355T 16362C 16399G 16519C	
890	A2	73G 146C 152C 153G 197G 235G 250G 263G 309.1C 315.1C 523d 524d 16111T 16223T 16290T 16319A 16362C 16519C	
891	L1b1a1'4	73G 152C 182T 185T 195C 247A 263G 315.1C 357G 523d 524d 16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C	
892	L3e2a1	73G 150T 195C 198T 263G 309.1CC 315.1C 16223T 16320T 16519C	
893	D1	73G 263G 315.1C 489C 16223T 16325C 16362C	
894	L3e2a1b1	73G 150T 195C 198T 263G 315.1C 499A 16223T 16320T 16399G 16519C	
895	L3b	73G 263G 315.1C 523d 524d 16124C 16189C 16223T 16278T 16362C 16519C	A16183C 3.71
896	Y2	73G 146C 204C 263G 309.1C 315.1C 482C 523d 524d 16126C 16231C 16311C	
898	A2d1a	64T 73G 146C 153G 235G 263G 315.1C 523d 524d 16111T 16223T 16274A 16290T 16319A 16362C	
900	L2a1+143+16189 (16192)+@16309	73G 143A 146C 152C 195C 263G 315.1C 16189C 16192T 16223T 16278T 16294T 16390A	C16221A 3.81 A16183C 2.27

Sample ID	Haplogroup	Haplotype	Heteroplasmy
901	C1b	73G 249d 263G 290d 291d 309.1C 315.1C 489C 493G 523d 524d 16131C 16223T 16291T 16298C 16325C 16327T	
903	A2ah	73G 146C 153G 235G 263G 309.1C 315.1C 523d 524d 16097C 16098G 16111T 16189C 16223T 16290T 16319A 16320T 16362C	
905	G1a2'3	73G 150T 263G 315.1C 325T 489C 16184T 16223T 16291T 16325C 16362C	
906	L2a1c1	73G 146C 152C 195C 198T 234G 263G 315.1C 16086C 16223T 16278T 16294T 16309G 16390A	A234A 46.31
907	D1h2	73G 152C 263G 309.1C 315.1C 489C 16223T 16239T 16260T 16311C 16325C 16362C	T195C 6.61
908	L3e1	73G 150T 189G 200G 263G 315.1C 524.1AC 16223T 16327T 16519C	A200A 4.64
909	L1c1	73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 315.1C 316A 523d 524d 16093C 16129A 16187T 16189C 16223T 16263C 16278T 16293G 16294T 16311C 16360T 16368C 16519C	T16093T 3.46
910	B4b	73G 263G 309.1CC 315.1C 499A 524.1ACAC 16182C 16183C 16189C 16193d 16217C 16519C	
911	L2a	73G 146C 152C 182T 263G 315.1C 316A 16086C 16114T 16129A 16169T 16223T 16239T 16240G 16274A 16278T 16291T 16390A 16519C	
914	L3e2b+152	73G 150T 152C 195C 263G 309.1CC 315.1C 16172C 16183C 16189C 16223T 16320T 16519C	
916	G2b2b	73G 146C 263G 315.1C 489C 16223T 16260T 16292T 16311C 16362C 16519C	
918	A2+(64)+16189	64T 73G 146C 153G 235G 263G 309.1CC 315.1C 523d 524d 16111T 16189C 16223T 16290T 16319A 16362C	
920	A2+(64)+16189	64T 73G 125C 127C 146C 153G 263G 309.1C 315.1C 499A 523d 524d 16111T 16183C 16189C 16223T 16290T 16319A 16362C 16519C	A16183A 13.44

Sample ID	Haplogroup	Haplotype	Heteroplasmy
922	L2a1b1a	73G 146C 152C 195C 263G 315.1C 16182C 16183C 16189C 16193d 16223T 16278T 16290T 16294T 16309G 16390A	
923	N9a1	73G 150T 195C 263G 309.1C 315.1C 16111T 16129A 16223T 16257A 16261T 16311C	
924	L3b1a3	73G 263G 315.1C 523d 524d 16124C 16223T 16278T 16311C 16362C 16519C	
925	L3e2a1	73G 150T 195C 198T 200G 263G 309.1C 315.1C 16148T 16223T 16320T 16497G 16519C	A200A 2.42
926	L2a5	73G 152C 182T 263G 315.1C 511T 16223T 16224C 16263C 16278T 16291T 16309G 16390A 16519C	
927	L4b1a	73G 150T 199C 204C 263G 309.1C 315.1C 513A 523d 524d 16182C 16183C 16189C 16193d 16223T 16239T 16311C 16320T 16362C 16519C	
928	A2h1	64T 73G 146C 153G 235G 263G 310C 315d 499A 523d 524d 16111T 16223T 16290T 16319A 16335G 16362C 16526A	
930	B2d	73G 263G 309.1C 315.1C 498d 499A 16183C 16189C 16217C 16519C	
931	R9b1b	58C 73G 152C 263G 309.1C 315.1C 16124C 16148T 16304C 16309G 16327T 16390A 16519C	A16309A 2.16
932	L2a1+16189 (16192)	73G 146C 152C 195C 263G 309.1C 315.1C 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C	C16192C 11.45 C16179T 4.69
934	B4a2b	73G 152C 263G 309.1C 315.1C 498d 499A 16183C 16189C 16217C 16360T 16519C	
935	A2+(64)+@16111	64T 73G 146C 153G 195C 235G 263G 309.1C 315.1C 513A 523d 524d 16223T 16290T 16319A 16362C	
936	L3f1b1a	73G 189G 200G 263G 309.1C 315.1C 16129A 16209C 16223T 16292T 16295T 16311C 16368C 16519C	
937	A5a	73G 235G 263G 315.1C 523d 524d 16187T 16223T 16290T 16319A	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
939	L1c1c	73G 151T 152C 182T 186A 189C 195C 247A 249d 263G 297G 315.1C 316A 523d 524d 16129A 16169T 16172C 16187T 16189C 16223T 16239T 16278T 16291T 16311C 16360T 16519C 16527T	
940	M71a2	73G 143A 146C 151T 189G 263G 315.1C 489C 16129A 16140C 16223T 16271C 16519C	A189A 38.05 G143G 2.75
941	D4e1	73G 263G 309.1C 315.1C 489C 16092C 16223T 16362C 16519C	
942	B4c1b2c2	73G 146C 150T 195C 263G 309.1C 315.1C 16129A 16138G 16140C 16166G 16183C 16184A 16189C 16193.1C 16217C 16274A 16335G 16519C	
943	L1b1a18	73G 152C 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523d 524d 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C	
944	L3d5	73G 152C 195C 199C 263G 315.1C 523d 524d 16124C 16223T 16362C 16519C	G16274A 6.31
945	A2	73G 146C 153G 235G 263G 309.1C 315.1C 523d 524d 16111T 16223T 16290T 16319A 16362C	
947	L2a1+16189 (16192)	73G 146C 152C 195C 263G 309.1CC 315.1C 16093C 16134T 16189C 16192T 16201T 16223T 16278T 16294T 16309G 16390A	T16093T 4.12
948	M7b1a1b	73G 150T 199C 204C 263G 271T 309.1C 315.1C 489C 16223T 16297C	
951	Z+152	73G 152C 214G 249d 263G 315.1C 489C 16185T 16223T 16260T 16298C	
956	L3e1a1	73G 150T 152C 185A 189G 194T 200G 263G 315.1C 16157C 16185T 16223T 16311C 16327T	C16185C 4.61
958	N9a	73G 150T 263G 309.1C 315.1C 16223T 16257A 16261T 16293G	
959	D6a2	73G 263G 315.1C 489C 573.1C 16129A 16223T 16274A 16311C 16317G 16362C 16519C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
960	B4b	73G 115C 263G 315.1C 499A 523d 524d 16183C 16189C 16217C 16296T 16311C 16519C	A200G 19.64
961	A2f2	64T 73G 146C 153G 235G 263G 309.1C 315.1C 368G 523d 524d 16111T 16223T 16311C 16319A 16357C 16362C 16519C	T16093C 10.06
964	M8a2a	73G 263G 309.1C 315.1C 489C 16184T 16189C 16223T 16298C 16319A	C16184C 2.02 A189G 3.41
965	D4b2a1	73G 263G 280T 315.1C 489C 523d 524d 16223T 16355T 16362C	
966	F2+16291	73G 249d 263G 315.1C 16037G 16189C 16291T 16304C 16519C	
967	L2c	73G 89C 93G 146C 150T 152C 182T 195C 263G 309.1C 315.1C 325T 523d 524d 16192T 16223T 16261T 16278T 16390A	
968	D4a	73G 263G 309.1C 315.1C 489C 16093C 16129A 16223T 16295T 16362C	T16093T 3.05
969	L1b1a7a	73G 152C 182T 185T 195C 228A 247A 263G 315.1C 357G 523d 524d 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C	A189G 3.48
971	L2a1a1	73G 146C 151T 152C 195C 263G 315.1C 16223T 16278T 16294T 16309G 16368C 16390A 16519C	
972	L3e2b	73G 150T 195C 263G 309.1C 315.1C 16172C 16183C 16189C 16223T 16261T 16284G 16320T 16519C	A16183A 2.56
973	L3b	73G 263G 315.1C 523d 524d 16093C 16124C 16223T 16278T 16362C 16519C	T16189C 33.18 T16093T 15.97
974	C1b	73G 198T 249d 263G 290d 291d 309.1CC 315.1C 489C 493G 523d 524d 16223T 16298C 16325C 16327T	
975	L3f1b1a	73G 189G 200G 263G 309.1C 315.1C 16129A 16209C 16223T 16292T 16295T 16311C 16519C	
976	A2ac	64T 73G 146C 153G 235G 263G 309.1C 315.1C 523d 524d 16111T 16213A 16223T 16290T 16319A 16362C	
978	L3e2b	73G 150T 195C 263G 315.1C 16172C 16183C 16189C 16223T 16320T 16519C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
979	L3f1b1	73G 150T 189G 200G 263G 315.1C 16209C 16223T 16292T 16295T 16311C 16519C 16527T	
980	L2a1+16189 (16192)	73G 146C 152C 195C 263G 309.1C 315.1C 16189C 16192T 16223T 16278T 16294T 16309G 16390A	
982	L3e2b+152	73G 150T 152C 195C 263G 297G 309.1C 315.1C 16147T 16172C 16183C 16188T 16189C 16193.1C 16223T 16320T 16519C	T16046C 5.96
983	L1c3b2	73G 151T 152C 182T 186A 189C 247A 263G 315.1C 316A 523d 524d 16086C 16129A 16163G 16187T 16189C 16223T 16278T 16293G 16294T 16311C 16360T 16519C	
984	L2a1a2	73G 146C 152C 195C 263G 315.1C 16223T 16278T 16286T 16294T 16309G 16390A 16519C	
985	M7c1a4a	73G 146C 199C 263G 309.1CC 315.1C 489C 523d 524d 16223T 16295T 16362C 16519C	
986	L2a1+143+@16309	73G 143A 146C 152C 195C 263G 315.1C 16129A 16223T 16278T 16294T 16390A	
988	L2a1+16189 (16192)	73G 146C 152C 195C 263G 315.1C 16189C 16192T 16223T 16278T 16294T 16309G 16390A	G16274A 7.17
989	L3b1a+@16124	73G 263G 309.1C 315.1C 484G 523d 524d 16223T 16278T 16355T 16362C 16519C	
990	L1c2a1a	73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 315.1C 316A 16071T 16129A 16145A 16187T 16189C 16213A 16223T 16234T 16265C 16278T 16286G 16294T 16311C 16360T 16527T	C16223C 3.18 G16213G 3.38 C16234C 5.78
992	B2d	73G 263G 309.1CC 315.1C 498d 499A 16183C 16189C 16217C 16519C	
993	A2+(64)+16129	64T 73G 146C 153G 235G 263G 309.1C 315.1C 523d 524d 16111T 16129A 16223T 16290T 16319A 16362C	
994	L2a1i	73G 143A 146C 152C 195C 263G 315.1C 16051G 16189C 16192T 16223T 16278T 16294T 16362C 16390A 16519C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
997	B4b	73G 263G 315.1C 461T 499A 16183C 16189C 16217C 16519C	
998	L2a1a2	73G 146C 152C 195C 263G 309.1C 315.1C 16223T 16278T 16286T 16294T 16309G 16390A 16519C	
999	C1	73G 249d 263G 290d 291d 309.1C 315.1C 489C 16223T 16298C 16325C 16327T	T16189C 4.97
1000	L3d1b3	73G 146C 152C 263G 309.1C 315.1C 523d 524d 16124C 16223T 16519C	
1002	L2c1a	73G 146C 150T 152C 182T 195C 198T 263G 309.1C 315.1C 325T 523d 524d 16147T 16223T 16261T 16278T 16318G 16390A	C16301T 2.72
1003	L3f1b+16292+150	73G 150T 189G 200G 263G 309.1C 315.1C 16209C 16223T 16292T 16311C 16519C	
1004	L2a1+16189 (16192)	73G 146C 152C 195C 263G 315.1C 16093C 16172C 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16362C 16390A	T16093T 3.08
1006	B2d	73G 263G 309.1C 315.1C 498d 499A 16183C 16189C 16217C 16519C	
1008	L3f1b4c	73G 150T 189G 200G 263G 309.1C 315.1C 16209C 16218T 16223T 16292T 16311C 16519C	
1009	L2a1+143+@16309	73G 143A 146C 152C 195C 263G 315.1C 16129A 16223T 16278T 16294T 16390A	
1011	B4	73G 146C 263G 309.1CC 315.1C 316A 524.1ACAC 16172C 16183C 16189C 16217C 16319A 16519C	A215G 2.47
1012	L3e2b+152	73G 150T 152C 195C 263G 309.1C 315.1C 16172C 16183C 16189C 16223T 16311C 16320T 16519C	
1014	B4b	73G 263G 315.1C 499A 16182C 16183C 16189C 16193d 16217C	
1015	L1b	73G 152C 182T 185T 195C 247A 263G 294C 315.1C 357G 523d 524d 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
1016	B4b	73G 263G 315.1C 499A 16178C 16183C 16189C 16217C 16519C	
1017	L2b1a	73G 150T 152C 182T 195C 198T 204C 207A 263G 315.1C 418T 523d 524d 16114A 16129A 16213A 16223T 16278T 16355T 16362C 16390A	
1018	L3f2a	73G 150T 152C 263G 315.1C 523d 524d 16209C 16223T 16290T 16311C 16319A 16362C 16519C	
1019	L2c3	73G 89C 93G 146C 150T 182T 195C 198T 263G 309.1C 315.1C 325T 513A 523d 524d 16223T 16257T 16278T 16390A	T204C 5.88
1020	A2ac	64T 73G 146C 235G 263G 309.1C 315.1C 523d 524d 16111T 16213A 16223T 16290T 16319A	
1021	W+194	73G 143A 189G 194T 195C 204C 207A 263G 315.1C 16181G 16223T 16292T 16519C	C194C 6.68
1022	L0a1a2	64T 93G 185A 189G 200G 247A 263G 315.1C 514T 523d 524d 16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16362C	
1023	M7a+16324	73G 150T 263G 309.1C 315.1C 489C 523d 524d 16209C 16223T 16290T 16324C	
1024	C1b	73G 199C 249d 263G 290d 291d 315.1C 489C 493G 523d 524d 16223T 16298C 16325C 16327T	
1026	G2b2c	73G 263G 309.1C 315.1C 489C 16183G 16189C 16223T 16264T 16355T 16362C	C16355C 3.61
1028	F3b	73G 249d 263G 309.1C 315.1C 16183C 16189C 16220C 16254G 16298C 16362C	
1029	L1b1a18	73G 152C 185T 189G 195C 247A 263G 315.1C 357G 523d 524d 16126C 16172C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C	
1031	L0a2a2	64T 93G 152C 189G 204C 207A 236C 247A 263G 315.1C 523d 524d 16148T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
1032	A2+(64)	64T 73G 146C 153G 235G 263G 309.1CC 315.1C 523d 524d 16111T 16223T 16290T 16319A 16362C 16518A 16519C	
1033	L3e2b	73G 150T 195C 263G 309.1C 315.1C 16172C 16183C 16188T 16189C 16193.1C 16223T 16320T 16519C	
1034	A2+(64)+@153	9A 64T 73G 146C 235G 263G 309.1CC 315.1C 523d 524d 16111T 16124C 16223T 16290T 16319A 16362C 16519C	
1035	L3f1b4c	73G 150T 189G 200G 263G 315.1C 385G 16188T 16209C 16218T 16223T 16292T 16311C 16519C	T204C 8.38
1039	D1i2	73G 263G 309.1C 315.1C 417A 489C 551G 16223T 16274A 16325C 16362C 16368C	A215G 4.62
1040	L2a1+143+16189 (16192)+@16309	73G 143A 146C 152C 195C 263G 309.1CC 315.1C 533G 16189C 16192T 16223T 16278T 16294T 16390A	
1041	M9	73G 263G 309.1C 315.1C 489C 523d 524d 16223T 16362C	
1042	L3b	73G 263G 315.1C 523d 524d 16124C 16223T 16278T 16362C 16519C	
1043	L3e2b+152	73G 150T 152C 195C 263G 315.1C 16172C 16183C 16189C 16223T 16320T 16519C	
1044	F1a4a	73G 152C 249d 263G 309.1CC 315.1C 521d 522d 523d 524d 16129A 16172C 16222T 16304C 16311C 16362C 16519C	
1046	L1c2	73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 309.1C 315.1C 316A 523G 16129A 16187T 16189C 16223T 16265C 16274A 16278T 16286G 16294T 16311C 16360T 16519C 16527T	
1048	D1	73G 263G 315.1C 489C 16223T 16325C 16362C	
1049	M9	73G 263G 309.1CC 315.1C 489C 16223T 16362C	
1050	L3e1	73G 150T 152C 189G 263G 309.1C 315.1C 16192T 16223T 16326G 16327T	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
1052	A2+(64)	64T 73G 146C 153G 235G 263G 309.1C 315.1C 523d 524d 16111T 16223T 16290T 16319A 16362C	
1055	L2a1	73G 146C 152C 195C 263G 315.1C 16086C 16189C 16223T 16278T 16294T 16309G 16390A	
1057	L0a1b2	93G 95C 185A 189G 236C 247A 263G 315.1C 523d 524d 16093C 16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T	T16093T 9.75
1058	M75	73G 146C 150T 152C 263G 310C 315d 489C 523d 524d 16068C 16183C 16189C 16223T 16311C 16519C	
1060	R9c1b1	73G 151T 263G 309.1CC 315.1C 479G 16157C 16304C	
1061	L1b1a+189	73G 152C 182T 185T 189G 195C 247A 263G 315.1C 357G 523d 524d 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C	
1066	L3b	73G 263G 315.1C 523d 524d 16124C 16223T 16278T 16362C 16519C	
1067	A2	73G 146C 152C 153G 235G 263G 309.1C 315.1C 523d 524d 16111T 16223T 16290T 16319A 16362C 16519C	
1068	L4b1a	73G 150T 199C 204C 207A 263G 309.1C 315.1C 513A 16179T 16182C 16183C 16189C 16223T 16239T 16311C 16320T 16362C 16519C	A16182A 2.13
1069	L3e2b+152	73G 150T 152C 195C 263G 315.1C 16172C 16183C 16189C 16223T 16263C 16320T 16519C	
1073	L3e2a1	73G 150T 195C 198T 263G 309.1C 315.1C 16223T 16320T 16519C	
1075	A2+(64)	64T 73G 146C 153G 235G 263G 309.1C 315.1C 523d 524d 16111T 16223T 16290T 16319A 16362C 16519C 16527T	
1076	B5a	73G 210G 263G 315.1C 523d 524d 16140C 16182C 16183C 16189C 16193d 16266A 16519C	
1077	L3e3	73G 150T 195C 263G 315.1C 523d 524d 16093C 16183C 16189C 16223T 16265T	A16183A 2.07 T16093T 2.28

Sample ID	Haplogroup	Haplotype	Heteroplasmy
1078	Y2	73G 153G 263G 315.1C 482C 16126C 16231C 16311C 16362C	T195C 4.15
1081	L3e2b	73G 150T 195C 263G 315.1C 16172C 16183C 16189C 16223T 16320T 16519C	
1082	L2a1a3c	73G 143A 146C 152C 195C 263G 315.1C 16093C 16223T 16256T 16278T 16294T 16309G 16390A 16519C	T16093T 3.92
1083	B4b1a+207	73G 143A 207A 263G 309.1CC 315.1C 499A 16136C 16182C 16183C 16189C 16193d 16217C 16234T 16519C	
1085	M7b1a1+(16192)	73G 150T 199C 263G 309.1C 315.1C 489C 16129A 16192T 16223T 16297C	
1086	B4	73G 263G 309.1CC 315.1C 16182C 16183C 16188T 16189C 16217C 16519C	C16188C 2.46 T16189T 2.01
1087	A2	73G 146C 153G 195C 235G 263G 309.1C 315.1C 523d 524d 16111T 16223T 16290T 16319A 16362C 16468C	
1088	L1c2a1a	73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 309.1C 315.1C 316A 16071T 16129A 16145A 16187T 16189C 16213A 16223T 16234T 16248T 16265C 16278T 16286G 16294T 16311C 16360T	
1089	L3e2a1	26T 55C 57C 73G 150T 195C 198T 263G 315.1C 16124C 16223T 16320T 16519C	A183G 7.88
1090	M7c1b2a	73G 146C 152C 199C 263G 315.1C 489C 523d 524d 16172C 16223T 16295T 16519C	
1092	L1c3b2	73G 151T 152C 182T 186A 189C 247A 263G 315.1C 316A 523d 524d 16086C 16104T 16163G 16187T 16189C 16223T 16278T 16293G 16294T 16311C 16360T 16519C	
1094	L3b2	73G 263G 315.1C 523d 524d 16124C 16223T 16362C 16527T	
1095	M8a2+152	73G 152C 263G 309.1CC 315.1C 489C 16184T 16223T 16293C 16298C 16319A 16519C	
1096	D4q	73G 200G 263G 309.1C 315.1C 489C 16223T 16256T 16264T 16311C 16362C 16519C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
1097	L2a1+143+@16309	73G 146C 152C 195C 263G 309.1C 315.1C 16086C 16183C 16189C 16223T 16278T 16294T 16390A 16519C	A16182C 26.56
1098	L3b3	73G 150T 185A 189G 263G 315.1C 523d 524d 16048A 16124C 16223T 16278T 16362C	
1099	L3b	73G 263G 315.1C 523d 524d 16124C 16223T 16278T 16362C 16519C	
1100	A14	73G 151T 152C 200G 235G 263G 315.1C 523d 524d 16176T 16223T 16290T 16319A 16362C 16463G 16526A	
1101	L3b1a+@16124	73G 263G 315.1C 523d 524d 16223T 16278T 16362C 16527T	
1103	B2+16278	73G 263G 315.1C 337T 499A 16183C 16189C 16217C 16278T 16311C 16519C	
1104	L3e2a1b2	73G 150T 195C 198T 263G 315.1C 16223T 16311C 16320T 16519C	
1106	L2a1i	73G 143A 146C 152C 195C 263G 315.1C 16189C 16192T 16223T 16278T 16294T 16362C 16390A 16519C	
1108	L3e2	73G 150T 195C 263G 315.1C 16223T 16320T 16519C	
1109	L2b1a	73G 150T 152C 182T 195C 198T 204C 263G 315.1C 418T 523d 524d 16114A 16129A 16213A 16223T 16278T 16355T 16362C 16390A	
1110	B2g	73G 114G 263G 315.1C 499A 16223T 16234T 16519C	C114C 43.27
1111	L3e2b+152	73G 150T 152C 195C 263G 309.1C 315.1C 16172C 16183C 16189C 16223T 16320T 16519C	A16183A 2.12
1112	D4e1a	73G 94A 263G 315.1C 489C 16092C 16223T 16362C	
1113	L2a1l	73G 143A 146C 152C 195C 263G 315.1C 534T 16189C 16192T 16223T 16235G 16278T 16294T 16309G 16390A	C16192C 2.2
1115	L2a1	73G 146C 152C 195C 263G 309.1C 315.1C 16223T 16278T 16294T 16309G 16390A 16519C	
1118	A+152+16362+200	73G 152C 200G 235G 263G 315.1C 523d 524d 16172C 16223T 16290T 16319A 16362C 16519C	G16390A 2.79

Sample ID	Haplogroup	Haplotype	Heteroplasmy
1120	A15	73G 152C 204C 207A 235G 309.1C 315.1C 523d 524d 16223T 16290T 16319A 16362C	
1121	G1a1	73G 146C 150T 263G 315.1C 489C 16223T 16325C 16362C 16468C 16519C	A183G 9.61
1123	L1b1a+189	73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523d 524d 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 16527T	A16293A 7.62
1124	L1b1a+189	73G 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523d 524d 16126C 16187T 16189C 16223T 16270T 16278T 16293G 16311C 16368C 16519C	
1125	A2z	73G 146C 152C 153G 214G 235G 263G 309.1C 315.1C 523d 524d 16083T 16111T 16223T 16256T 16290T 16319A 16362C	C16360T 28.06
1126	L1c1a+@198	73G 151T 152C 182T 186A 189C 195C 247A 263G 297G 315.1C 316A 523d 524d 16093C 16129A 16187T 16189C 16223T 16263C 16278T 16293G 16294T 16311C 16360T 16368C 16519C	T16093T 2.81
1127	L3d	73G 152C 263G 309.1C 315.1C 523d 524d 16124C 16223T 16291T	
1128	A2	73G 146C 153G 235G 263G 315.1C 523d 524d 16111T 16223T 16290T 16319A 16362C	
1130	F3a1	73G 249d 263G 309.1C 315.1C 16093C 16260T 16298C 16355T	T16093T 3.54
1131	B5a	73G 152C 210G 263G 310C 315d 523d 524d 16140C 16183C 16189C 16266A 16519C	A16559G 3.19
1132	L2c2	73G 89C 93G 146C 150T 152C 182T 195C 198T 263G 309.1C 315.1C 325T 523d 524d 16183G 16214T 16223T 16264T 16278T 16390A	
1134	M7b1a1+(16192)	64T 73G 150T 199C 263G 315.1C 489C 16129A 16192T 16223T 16297C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
1135	L2c2	73G 93G 146C 150T 152C 182T 195C 198T 263G 309.1CC 315.1C 325T 523d 524d 16223T 16264T 16278T 16291T 16390A	T16368C 4.6
1136	D4g2a	73G 263G 298T 309.1C 315.1C 489C 16223T 16274A 16325C 16362C	
1138	B2c2	73G 146C 203A 222T 263G 309.1C 315.1C 499A 16168T 16183d 16187T 16189C 16217C 16355T 16508T 16512C 16519C 16520T	T16519T 2.31
1140	L3e2b	73G 150T 195C 263G 315.1C 16172C 16183C 16189C 16223T 16242T 16320T 16519C	T16172T 2.09
1142	W1+119	73G 119C 189G 195C 204C 207A 263G 315.1C 16223T 16292T 16295T 16311C 16519C	T152C 18.34
1143	L3d1a1a	73G 150T 152C 263G 315.1C 523d 524d 16124C 16223T 16319A	
1144	L3b1a3	73G 263G 315.1C 523d 524d 16093C 16117C 16124C 16223T 16278T 16311C 16362C 16519C	T16093T 6.22
1145	L3d1b2	73G 150T 152C 263G 315.1C 523d 524d 573.1C 16124C 16223T 16288C 16519C	
1148	L3d	73G 263G 315.1C 523d 524d 16124C 16223T	
1153	D1i2	73G 263G 315.1C 417A 489C 551G 16223T 16274A 16325C 16362C 16368C	
1155	M7b1a1+(16192)	73G 150T 182T 199C 263G 315.1C 459d 489C 16129A 16192T 16223T 16297C 16519C	T204C 3.71
1156	L3h1a2b	73G 263G 315.1C 16093C 16223T 16270T 16311C	T16093T 5.07
1157	L3f	73G 228A 263G 315.1C 16126C 16209C 16223T 16519C	
1158	L2a1a1	73G 146C 152C 195C 263G 309.1CC 315.1C 524.1AC 16129A 16223T 16278T 16294T 16309G 16368C 16390A 16519C	C16294C 4.48
1159	A2	73G 146C 153G 235G 263G 309.1CC 315.1C 523d 524d 16111T 16223T 16290T 16319A 16362C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
1160	L1c2	73G 151T 152C 182T 186A 189C 195C 198T 204C 247A 263G 297G 315.1C 316A 513A 16129A 16187T 16189C 16214T 16223T 16265C 16278T 16286A 16291T 16294T 16311C 16360T 16519C 16527T	
1162	L1c1d	73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 309.1C 315.1C 316A 523d 524d 16038G 16086C 16129A 16187T 16189C 16223T 16278T 16284G 16293G 16294T 16311C 16360T 16519C	T204C 4
1164	L2a1a1	73G 146C 152C 195C 263G 309.1C 315.1C 524.1ACAC 16223T 16278T 16294T 16309G 16368C 16390A 16519C	
1165	L2a1a3	73G 143A 152C 195C 263G 309.1C 315.1C 16093C 16223T 16278T 16294T 16309G 16390A 16519C	T16093T 3.88
1166	L1c3c	73G 151T 152C 182T 186A 189C 195C 247A 248G 263G 315.1C 316A 523d 524d 16129A 16187T 16189C 16223T 16278T 16293G 16294T 16311C 16360T 16519C	A215G 28.97
1168	B2c2	73G 114T 146C 152C 263G 309.1CC 315.1C 499A 16086C 16182C 16183C 16189C 16193d 16217C 16519C	
1170	L3f1b+16292+150	73G 150T 189G 200G 204C 263G 309.1C 315.1C 16209C 16223T 16292T 16311C 16519C	T204T 46.54
1171	L2a1+143	73G 143A 146C 152C 189G 195C 263G 309.1C 315.1C 523d 524d 16124C 16223T 16278T 16294T 16309G 16390A	
1172	L2a1c+16129	73G 143A 146C 152C 195C 263G 315.1C 16037G 16129A 16223T 16278T 16294T 16309G 16390A	T139C 3.27
1173	M7b1a1+(16192)	73G 150T 199C 263G 315.1C 489C 16129A 16192T 16223T 16297C	
1174	L3e1	73G 150T 152C 189G 200G 263G 315.1C 523d 524d 16223T 16327T	A215G 10.79
1176	B4	73G 195C 263G 309.1CC 315.1T 524.1AC 16182C 16183C 16189C 16193d 16217C 16310A 16519C	A16182A 2.43 A16183A 2.55

Sample ID	Haplogroup	Haplotype	Heteroplasmy
1177	L2a1+143	73G 143A 146C 152C 195C 263G 309.1C 315.1C 16278T 16294T 16309G 16355T 16390A	
1180	F1a1c	73G 200G 249d 263G 315.1C 523d 524d 548T 16129A 16162G 16221T 16304C 16519C	
1181	L1b1a10	73G 151T 152C 182T 185T 189G 195C 247A 263G 315.1C 357G 523d 524d 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16311C 16519C	
1182	E1a1a1	73G 263G 315.1C 456T 489C 16223T 16291T 16362C 16390A 16519C	
1183	L1b	73G 152C 182T 185T 195C 198T 247A 263G 315.1C 357G 523d 524d 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16290T 16293G 16311C 16519C	
1184	B4b	73G 140T 263G 309.1CC 315.1C 499A 16182C 16183C 16189C 16193d 16217C 16301T 16304C 16519C	
1185	D4h3a	73G 152C 263G 315.1C 489C 16223T 16241G 16301T 16342C 16362C	
1187	L3d1d	73G 152C 263G 315.1C 523d 524d 16124C 16223T 16256T 16368C	
1188	L3e2b+152	73G 152C 195C 263G 315.1C 16172C 16183C 16189C 16223T 16320T 16519C	
1189	L1c2	73G 151T 152C 182T 186A 189C 195C 247A 263G 297G 315.1C 316A 16129A 16187T 16189C 16223T 16265C 16278T 16286G 16294T 16311C 16359C 16360T 16519C 16527T	
1190	L1c2	73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 309.1C 315.1C 316A 16129A 16187T 16189C 16223T 16265C 16278T 16286G 16288C 16294T 16311C 16360T 16519C 16527T	
1191	A2+(64)	64T 73G 146C 153G 235G 263G 309.1C 315.1C 523d 524d 16111T 16223T 16290T 16319A 16362C	
1192	L0a1'4	93G 185A 189G 236C 247A 263G 315.1C 523d 524d 16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
1193	B4c1a1	73G 263G 309.1CC 315.1C 16183C 16189C 16217C 16311C 16519C	
1194	G2a1+16189	73G 263G 309.1C 315.1C 489C 16189C 16223T 16227G 16278T 16362C	
1196	L3b1a+@16124	73G 263G 315.1C 523d 524d 16093C 16223T 16278T 16362C 16519C	T16093T 3.88
1197	L0a1b1	93G 95C 185A 189G 236C 247A 263G 315.1C 523d 524d 16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T	A16230A 6.85 C16223C 7.97
1198	L1c3a	73G 151T 152C 182T 186A 189C 247A 263G 315.1C 316A 523d 524d 16129A 16145A 16189C 16215G 16223T 16278T 16294T 16311C 16360T 16519C	A16183C 2.11
1200	M7c1a1b	73G 199C 263G 309.1C 315.1C 489C 523d 524d 16092C 16223T 16295T 16304C 16311C 16519C	T16092T 2.64
1201	R9b1b	58C 73G 152C 263G 309.1C 315.1C 16124C 16148T 16304C 16309G 16327T 16390A 16519C	
1202	L1c3a	73G 151T 152C 182T 186A 189C 194T 247A 263G 315.1C 316A 523d 524d 16129A 16189C 16215G 16223T 16278T 16294T 16311C 16360T 16519C	A16183C 5.67 C16292T 16.74
1203	A2+(64)	64T 73G 146C 153G 235G 263G 309.1C 315.1C 523d 524d 16104T 16111T 16223T 16290T 16294T 16319A 16362C 16519C	
1204	L3d1c	64T 73G 152C 263G 295T 309.1C 315.1C 523d 524d 16124C 16166G 16189C 16223T 16320T	
1205	L3f1b1	73G 189G 263G 315.1C 16183C 16189C 16209C 16223T 16292T 16295T 16311C 16519C	A215G 11.85
1206	L3f1b1	73G 189G 263G 315.1C 16183C 16189C 16209C 16223T 16292T 16295T 16311C 16519C	A215G 6.58
1207	L1b1a+189	73G 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523d 524d 16126C 16187T 16189C 16213A 16223T 16264T 16270T 16278T 16311C 16519C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
1209	L3f1b+16292+150	73G 150T 189G 200G 215G 263G 309.1C 315.1C 16209C 16223T 16292T 16311C 16519C	
1210	C1c1	73G 215G 249d 263G 290d 291d 309.1CC 315.1C 489C 503T 16223T 16298C 16325C 16327T	A215A 28.92
1213	L1b1a1'4	73G 152C 182T 185T 195C 247A 263G 315.1C 357G 523d 524d 16114A 16126C 16187T 16189C 16223T 16234T 16239T 16264T 16270T 16278T 16293G 16311C 16519C	
1214	L0a1a2	64T 93G 185A 189G 200G 236C 247A 263G 315.1C 523d 524d 16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C	
1216	C1b	73G 150T 249d 263G 290d 291d 309.1C 315.1C 489C 493G 523d 524d 16092C 16223T 16274A 16298C 16325C 16327T 16519C	T16092T 3.59
1217	A2ad	64T 73G 153G 194T 235G 263G 315.1C 523d 524d 16111T 16126C 16175G 16290T 16300G 16319A 16362C 16519C	
1219	D6a1	73G 263G 315.1C 489C 16182C 16183C 16189C 16223T 16274A 16362C	
1224	L2a1a2	73G 146C 152C 195C 263G 309.1C 315.1C 16136C 16189C 16223T 16278T 16286T 16294T 16309G 16390A 16519C 16524T	A16524A 39.63
1226	L2c	73G 93G 146C 150T 152C 182T 195C 198T 263G 315.1C 325T 523d 524d 16223T 16278T 16390A 16519C	
1227	D4b2b	73G 194T 263G 315.1C 489C 523d 524d 16223T 16362C 16519C	
1228	L2a1	73G 146C 152C 195C 263G 315.1C 16223T 16278T 16294T 16309G 16390A 16519C	
1229	L3b	73G 263G 315.1C 523d 524d 16124C 16223T 16278T 16362C 16519C	
1230	L2a1l	73G 143A 146C 152C 195C 263G 315.1C 534T 16189C 16192T 16223T 16278T 16294T 16309G 16390A 16519C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
1232	L3h1a2a	73G 146C 152C 263G 315.1C 16148T 16223T 16292T 16311C 16399G	
1233	L3e2a1b1	73G 150T 195C 198T 263G 315.1C 499A 16223T 16242A 16320T 16399G 16519C	C16242C 9.55
1235	M9a1a1b	73G 153G 263G 315.1C 489C 16223T 16234T 16300G 16316G 16362C	T16243C 4.84
1237	E1a1a1	73G 263G 309.1C 315.1C 372C 489C 16221T 16223T 16291T 16362C 16390A 16519C	T152C 15.39
1239	C1b	73G 249d 263G 290d 291d 309.1C 315.1C 489C 493G 523d 524d 16298C 16325C 16327T 16519C	
1240	L3e1d	73G 150T 152C 189G 200G 263G 309.1C 315.1C 16176T 16223T 16327T	C568T 11.51
1242	L3e2b+152	73G 150T 152C 195C 263G 309.1C 315.1C 16172C 16189C 16223T 16320T 16519C	
1243	L3b1a8	73G 152C 263G 315.1C 523d 524d 16145A 16223T 16278T 16362C 16519C	
1244	L3e3	73G 150T 195C 263G 309.1C 315.1C 523d 524d 16223T 16265T 16519C	
1246	B5b1c	73G 103A 152C 204C 263G 309.1CC 315.1C 523d 524d 16140C 16189C 16243C 16519C	T199C 3.66
1251	L3f1b+16292	73G 189G 200G 263G 309.1C 315.1C 16129A 16209C 16223T 16292T 16311C 16519C	
1265	G2a	73G 195C 263G 272G 309.1C 315.1C 489C 16188T 16223T 16227G 16278T 16362C	A272A 2.45 T489T 2.27 C16256A 2.09
1272	M72	73G 263G 309.1C 315.1C 489C 16166d 16184T 16214T 16223T	A16166A 15.03 T489T 4.16
1286	L3e1e2	90A 97A 106d 107d 108d 109d 110d 111d 150T 189G 200G 263G 315.1C 16223T 16327T	
1287	D1	73G 263G 309.1CC 315.1C 489C 16183C 16189C 16223T 16325C 16362C	
1290	C4b1	73G 146C 249d 263G 309.1C 315.1C 489C 16223T 16298C 16327T 16519C	

Sample ID	Haplogroup	Haplotype	Heteroplasmy
1291	D5c	73G 151T 152C 263G 309.1C 315.1C 489C 16079T 16145A 16189C 16190T 16193.1CC 16223T 16362C	C16190C 2.95 T489T 2.03 T16189T 2.67
1294	B4b1b'c	73G 263G 309.1CC 315.1C 499A 16136C 16183C 16189C 16217C 16218T 16519C	
1303	C4a1a+195	73G 195C 249d 263G 266C 315.1C 489C 16093C 16129A 16223T 16298C 16327T 16519C	T16093T 5.44
1305	M7b1a1+(16192)	73G 150T 199C 263G 309.1C 315.1C 332T 489C 16129A 16192T 16223T 16297C 16324C	C16256A 2.18
1308	L1c2	73G 152C 182T 186A 189C 195C 198T 247A 263G 297G 309.1C 315.1C 316A 16129A 16187T 16189C 16223T 16265C 16278T 16286A 16292T 16294T 16311C 16360T 16519C 16527T	
1309	A+152+16362	73G 152C 182T 235G 263G 315.1C 523d 524d 16223T 16248T 16290T 16319A 16357C 16362C	C16256A 6.8 T16249C 2.26
1310	M7c1a3	73G 146C 199C 263G 315.1C 489C 523d 524d 16295T 16319A 16519C	
1311	M9a1a1c	73G 146C 153G 263G 309.1CC 315.1C 489C 16154C 16223T 16234T 16291T 16316G 16362C	
1319	E1a1a1	73G 263G 309.1CC 315.1C 489C 16183C 16189C 16223T 16291T 16362C 16390A 16519C	

Supplementary Table 2: Observations of heteroplasmy by position per population group.

Nucleotide position	African, n=196	Asian, n=125	Latino, n=57	European, n=545	Total, n=923
46	0	0	0	1	1
64	0	0	0	2	2
66	0	0	1	2	3
114	0	1	0	2	3
119	0	0	0	2	2

Nucleotide position	African, n=196	Asian, n=125	Latino, n=57	European, n=545	Total, n=923
139	1	0	0	0	1
143	0	1	0	0	1
146	0	0	0	4	4
151	1	0	0	0	1
152	1	1	0	16	18
153	0	0	0	1	1
178	0	0	0	1	1
182	0	0	0	1	1
183	1	1	0	0	2
185	0	0	0	10	10
189	1	2	0	5	8
194	0	1	0	0	1
195	1	2	1	7	11
199	0	2	0	3	5
200	3	3	0	1	7
204	5	1	0	9	15
207	0	0	0	2	2
214	0	0	0	11	11
215	5	2	2	10	19
217	0	0	0	1	1
227	0	0	0	1	1
234	1	0	0	5	6
235	0	0	0	1	1
237	0	0	0	1	1
251	0	0	0	1	1
263	1	0	0	1	2
272	0	1	0	0	1
310	0	0	0	15	15
374	0	0	1	0	1
376	0	0	0	1	1
456	0	0	0	3	3
489	0	3	0	0	3
497	0	0	0	1	1
499	0	0	0	1	1
568	1	0	0	0	1
16037	0	0	0	1	1
16046	1	0	0	0	1

Nucleotide position	African, n=196	Asian, n=125	Latino, n=57	European, n=545	Total, n=923
16051	0	0	0	2	2
16069	0	0	0	2	2
16086	0	0	0	2	2
16092	0	2	1	4	7
16093	17	5	2	35	59
16126	0	0	0	1	1
16129	0	1	0	6	7
16131	0	1	0	0	1
16145	0	0	1	1	2
16148	0	1	0	1	2
16166	0	1	0	0	1
16169	0	0	0	1	1
16172	1	0	0	0	1
16179	1	1	0	1	3
16182	2	1	0	4	7
16183	10	1	2	25	38
16184	0	1	0	0	1
16185	1	0	0	0	1
16187	0	0	1	0	1
16188	0	0	1	0	1
16189	1	3	1	14	19
16190	0	1	0	0	1
16192	2	0	0	4	6
16194	0	1	0	0	1
16201	0	0	0	2	2
16209	2	0	0	0	2
16213	1	0	0	0	1
16217	0	0	1	0	1
16221	1	0	0	1	2
16222	0	0	0	1	1
16223	3	0	1	2	6
16227	0	1	0	0	1
16230	2	0	0	0	2
16234	2	1	1	4	8
16239	0	0	0	1	1
16241	0	0	0	1	1
16242	1	0	0	0	1

Nucleotide position	African, n=196	Asian, n=125	Latino, n=57	European, n=545	Total, n=923
16243	0	1	0	0	1
16249	0	1	0	0	1
16256	0	3	0	0	3
16258	0	0	0	2	2
16261	0	0	0	1	1
16265	0	0	0	3	3
16274	2	0	0	0	2
16278	0	0	0	1	1
16280	0	0	0	1	1
16286	0	0	0	1	1
16288	0	0	0	1	1
16292	1	0	0	1	2
16293	1	0	0	1	2
16294	1	0	0	1	2
16296	0	0	0	4	4
16297	0	1	0	2	3
16298	0	0	0	1	1
16299	0	0	0	1	1
16301	1	0	0	1	2
16304	0	0	0	4	4
16309	0	1	0	1	2
16311	0	0	0	4	4
16327	0	1	0	0	1
16335	0	0	0	3	3
16355	0	1	0	0	1
16356	0	0	0	1	1
16360	0	1	0	0	1
16362	0	0	0	1	1
16368	1	0	0	1	2
16390	0	1	0	2	3
16399	0	0	0	1	1
16488	0	1	0	0	1
16519	0	0	1	4	5
16524	1	0	0	1	2
16559	0	1	0	0	1
Sum	78	56	18	283	435

Supplementary Table 3: Haplogroup observations and population group category.

Population group	Haplogroup	Observations		Population group	Haplogroup	Observations
African	L3e2b+152	11		Asian	C1	1
African	L3e2b	10		Latino	C1b4	1
Latino	A2	7		Latino	C1c1	1
Asian	B4b	7		Latino	C1c4	1
African	L3b	7		Asian	C4a1a+195	1
Latino	A2+(64)	6		Asian	C4b1	1
African	L1c2	6		Latino	D1h2	1
African	L2a1	6		Asian	D4a3a2	1
Asian	M7b1a1+(16192)	6		Asian	D4a8	1
Latino	C1b	5		Asian	D4b2a1	1
Latino	D1	5		Asian	D4e1a	1
African	L1b	5		Asian	D4g2a	1
African	L1b1a+189	5		Latino	D4h3a	1
African	L2a1+16189 (16192)	5		Asian	D4q	1
African	L3b1a+@16124	5		Asian	D5b1c1	1
African	L3e3	5		Asian	D5c	1
Asian	B4	4		Asian	D6a1	1
African	L2a1+143+@16309	4		Asian	D6a2	1
African	L2a1a2	4		Asian	F1a	1
African	L2b	4		Asian	F1a1d	1
African	L3e1	4		Asian	F1a3+16311	1
African	L3e2a1	4		Asian	F1a4a	1
African	L3f1b1	4		Asian	F2+16291	1
Asian	A14	3		Asian	F3a1	1
Latino	B2c2	3		Asian	F3b	1
Latino	B2d	3		Asian	F4a1b	1
Asian	E1a1a1	3		Asian	G1a1	1
African	L0a1a2	3		Asian	G1a2'3	1
African	L1c1d	3		Asian	G2a	1
African	L1c2a1a	3		Asian	G2a1+16189	1
African	L1c3a	3		Asian	G2a1c	1

Population group	Haplogroup	Observations		Population group	Haplogroup	Observations
African	L2a1+143	3		Asian	G2b2b	1
African	L2a1a1	3		Asian	G2b2c	1
African	L2b1a	3		Asian	G2c	1
African	L2c2	3		Asian	G3	1
African	L3d1b2	3		African	L0a1'4	1
African	L3e1e2	3		African	L0a1b2	1
African	L3e2a1b1	3		African	L0d1'2	1
African	L3f1b+16292+150	3		African	L0d3b	1
African	L4b1a	3		African	L1b1a10	1
Asian	M9	3		African	L1b1a7a	1
Latino	A2+(64)+@153	2		African	L1c1	1
Latino	A2+(64)+16129	2		African	L1c1a+@198	1
Latino	A2+(64)+16189	2		African	L1c1a2b	1
Asian	A2ac	2		African	L1c2a1	1
Asian	B5a	2		African	L1c3b	1
Latino	C1b2	2		African	L1c3b'c	1
Latino	D1i2	2		African	L1c3c	1
Asian	D4a	2		African	L2a	1
Asian	D4b2b	2		African	L2a1a3	1
Asian	D4e1	2		African	L2a1a3c	1
Asian	D4g2	2		African	L2a1b1a	1
Asian	F1a1c	2		African	L2a1c+16129	1
African	L0a1b1	2		African	L2a1c1	1
African	L0a2a2	2		African	L2a5	1
African	L1b1a1'4	2		African	L2b2a	1
African	L1b1a18	2		African	L2c1a	1
African	L1c3b2	2		African	L3b1a8	1
African	L2a1+143+16189 (16192)+@16309	2		African	L3b2	1
African	L2a1i	2		African	L3b3	1
African	L2a1l	2		African	L3d1a1a	1
African	L2c	2		African	L3d1b3	1
African	L2c3	2		African	L3d1b3a	1
African	L3b1a3	2		African	L3d1c	1
African	L3d	2		African	L3d1d	1
African	L3d5	2		African	L3e1a1	1
African	L3e1a	2		African	L3e1b2	1

Population group	Haplogroup	Observations		Population group	Haplogroup	Observations
African	L3f1b1a	2		African	L3e1d	1
African	L3f1b4c	2		African	L3e1f1a	1
Asian	M7c1c2	2		African	L3e2	1
Asian	M9a1a1c	2		African	L3e2a1b2	1
Asian	N9a1	2		African	L3f	1
Asian	R9b1b	2		African	L3f1b+16292	1
Asian	Y2	2		African	L3f2a	1
Asian	A+152+16362	1		African	L3h1a2a	1
Asian	A+152+16362+200	1		African	L3h1a2b	1
Asian	A15	1		African	L3h1b1a	1
Latino	A2+(64)+@16111	1		African	L4b2b1	1
Asian	A2ad	1		Asian	M71a1a	1
Asian	A2ah	1		Asian	M71a2	1
Asian	A2ai	1		Asian	M72	1
Latino	A2d1a	1		Asian	M75	1
Latino	A2f2	1		Asian	M7a+16324	1
Latino	A2h	1		Asian	M7b1a1b	1
Latino	A2h1	1		Asian	M7c1	1
Latino	A2z	1		Asian	M7c1a1b	1
Asian	A5a	1		Asian	M7c1a3	1
Latino	B2+16278	1		Asian	M7c1a3a	1
Latino	B2g	1		Asian	M7c1a4a	1
Latino	B2h	1		Asian	M7c1b2a	1
Latino	B2o	1		Asian	M8a2+152	1
Asian	B4+16261	1		Asian	M8a2a	1
Asian	B4a1a	1		Asian	M9a1a1b	1
Asian	B4a2b	1		Asian	N9a	1
Asian	B4b1a+207	1		Asian	N9a1'3	1
Asian	B4b1b'c	1		Asian	P4b	1
Asian	B4c1a1	1		Asian	R9c1b1	1
Asian	B4c1b2c2	1		Asian	W+194	1
Asian	B5b1c	1		Asian	W1+119	1
Asian	B5b2+@204	1		Asian	Z+152	1
Asian	B5b2c	1		Asian	Z4	1

BIBLIOGRAPHY

Allard M, Wilson M, Monson K, Budowle B. Control region sequences for Far East individuals in the Scientific Working Group on DNA Analysis Methods forensic mtDNA data set. Legal Medicine. 2004:6:11-24.

Alves-Silva J, Silva Santos M, Guimarães P, Ferreira A, Bandelt H, Pena S, Prado, F.The ancestry of brazilian mtDNA lineages. American Journal of Human Genetics. 2000:67:444-461.

Anderson S, Bankier AT, Barrell BG, M.H. dB, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981;290:457-65.

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;23:147.

Bandelt HJ, Parson W. Consistent treatment of length variants in the human mtDNA control region: a reappraisal. International Journal of Legal Medicine. 2008;122(1):11-21.

Bryc K, Velez C, Karafet T, Moreno-Estrada A., Reynolds A., Auton A, Hammer M, Bustamante C, Ostrer H. Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(Suppl 2):8954-8961. Carracedo A, Bär W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland MM, Tully G, Wilson M. DNA commission of the International Society of Forensic Genetics: guidelines for mitochondrial DNA typing. Forensic Science International. 2000;110:79-85.

Chaitanya L, Ralf A, Oven M, Kupiec T, Chang J, Lagacé R, Kayser M. Simultaneous Whole Mitochondrial Genome Sequencing with Short Overlapping Amplicons Suitable for Degraded DNA Using the Ion Torrent Personal Genome Machine. Human Mutation. 2015;36: 1236–1247.

Chaitanya L, Oven M, Brauer S, Zimmermann B, Huber H, Xavier C, Parson W, 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 Sceince Interntional: Genetics. 2016;21:158-167.

Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. American Journal of Human Genetics. 1995;57(1):133-149.

Forster L, Forster P, Gurney SM, Spencer M, Huang C, Rohl A, Brinkmann, B. Evaluating length heteroplasmy in the human mitochondrial DNA control region. Int J Legal Med. 2010; 124(2):133–142.

Gallimore J, McElhoe J, Holland M. Assessing heteroplasmic variant drift in the mtDNA control region of human hairs using an MPS approach. Forensic Science International: Genetics. 2018:32:7-17.

Gentra Puregene Handbook: DNA Purification from a Buccal Brush using the Gentra Puregene Buccal Cell Kit. 3rd ed. Qiagen. 2011.

Gentra Puregene Handbook: DNA Purification from Body Fluid using the Gentra Puregene Buccal Cell Kit. 3rd ed. Qiagen. 2011

Gill P, Ivanov P, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K. Identification of the remains of the Romanov family by DNA analysis. Nature Genetics. 1994;6:130-135.

Greaves L, Nooteboom M, Elson J, Tuppen H, Taylor G, Commane D, Arasaradnam R, Khrapko K, Taylor R, Kirkwood T, Mathers J, Turnbull D. Clonal Expansion of Early to Mid-Life Mitochondrial DNA Point Mutations Drives Mitochondrial Dysfunction during Human Ageing. Public Library of Science: Genetics. 2014;10(9):e1004620.

Holland M, Parsons T. Mitochondrial DNA sequence analysis - Validation and use for forensic casework. Forensic Science Review. 1999;11:21-50.

Holland M, Pack E, McElhoe J. Evaluation of GeneMarker ® HTS for improved alignment of mtDNA MPS data, haplotype determination, and heteroplasmy assessment. Forensic Science International: Genetics. 2017:28:90-98.

Holland M, Makova K, McElhoe J. Deep-coverage MPS analysis of heteroplasmic variants withing the mtGenone allows for frequent differentiation of maternal relatives. Genes. 2018:9:124.

Irwin JA, Saunier JL, Niederstätter H, Strouss KM, Sturk KA, Diegoli TM, et al. 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.

Ivanov P, Wadhams M, Roby R, Holland M, Weedn V, Parsons T. 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;12:417-20.

Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, Villems R. The Emerging Limbs and Twigs of the East Asian mtDNA Tree. Molecular Biology and Evolution. 2002:19:1737-1751.

Kumar S, Bellis C, Zlojutro M, Melton P, Blangero J, Curran J. Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. BMC Evolutionary Biology. 2011:11:293.

McElhoe J, Holland M, Makova K, Su M, Paul I, Baker C, Faith S, Young B. Development and Assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. Forensic Science International: Genetics. 2014:13:20-29.

Melton T, Nelson K. Forensic mitochondrial DNA analysis: two years of commercial casework experience in the United States. Croat. Med. J. 2001:42:298-303.

STDEVA function. Microsoft web site. Availible at: https://support.office.com/en-us/article/stdeva-function-5ff38888-7ea5-48de-9a6d-11ed73b29e9d. Published: 2108. Accessed: July, 2018.

Parson W, Dür A. EMPOP- a forensic mtDNA database. Forensic Science International: Genetics. 2007:12:88-92.

Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Kashani BH, Ritchie KH, Scozzari R, Kong QP, Myres NM, Salas A, Semino O, Bandelt HJ, Woodward SR, Torroni A: Distinctive Paleo-Indian Migration Routes from Beringia Marked by Two Rare mtDNA Haplogroups. Current Biology. 2009:19: 1-8.

Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, Kashani B, Carossa V, Ekins J, Gómez-Carballa A, Huber G, Zimmermann B, Corach D, Babudri N, Panara F, Myres NM, Parson W, Semino O, Salas A, Woodward S, Achilli A, Torroni A, The initial peopling of the

Americas: A growing number of founding mitochondrial genomes from Beringia. Genome Research. 2010:20:1174–1179.

Rathbun M, McElhoe J, Parson W, Holland M. Considering DNA damage when interpreting mtDNA heteroplasmy in deep sequencing data. Forensic Science International: Genetics. 2017:26:1-11.

RStudio Team. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. 2016. URL http://www.rstudio.com/.

Salas A, Lareu M, Carracedo A. Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: a case report. International Journal of Legal Medicine. 2001;114:186-90.

Scientific Working Group on DNA Analysis Methods (SWGDAM). Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories. 2013.

Stoneking M, Hedgecock D, Higuchi R, Vigilant L, Ehrlich H. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence specific oligonucleotide probes. American Journal of Human Genetics. 1991:48:370-82.

Tanaka M, Cabrera V, González A, Larruga J, Takeyasu T, Fuku N, Guo L, Hirose R, Fujita Y, Kurata M. Mitochondrial Genome Variation in Eastern Asia and the Peopling of Japan. Genome Research. 2004:14:1832-1850

Umetsu K, Yuasa I. Recent progress in mitochondrial DNA analysis. Legal Medicine. 2005:7:259-262.

van Oven M. PhyloTree Build 17: Growing the human mitochondrial DNA tree. Forensic Science International: Genetics Supplement Series 2015:5: 9-11.

Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt HJ, Kronenberg F, Salas A, Schönherr S. HaploGrep2 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Research. 2016:44:58-63.

E.B., Wilson EB. Probable inference, the law of succession, and statistical inference. Journal of American Statistics Association. 1927;22:209-212.

Wilson M, DiZinno J, Polanskey D, Replogle J, Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. International Journal of Legal Medicine. 1995;108:68-74.

16S Metagenomic equencing Library Preparation (15044223 B). Rev. B. Illumina. 2013.

ACADEMIC VITA

EMILY L. DEMCHAK
emdemchak@yahoo.com

EDUCATION

Bachelor of Science in Forensic Science (Biology Option)
Schreyer's Honors College, The Eberly College of Science
The Pennsylvania State University, University Park, PA August 2018

PROFESSIONAL AND ACADEMIC EXPERIENCE

Undergraduate Research
Forensic Science Program, The Eberly College of Science
The Pennsylvania State University, University Park, PA Spring 2016 – Present

- Acquired knowledge in the realm of Massively Parallel DNA Sequencing (MiSeq Personal Sequencer, Illumina®, Inc.)
- Conducted eight sequencing runs following Nextera XT library preparation and SoftGenetics® GeneMarker® HTS data analysis workflow for the control region of mtDNA.
- Gained practical molecular biology skills including DNA extraction, Polymerase Chain Reaction amplification, DNA quantitation, Massively Parallel Sequencing data analysis, and sterile techniques required when working with human samples.
- Managed and tracked purchasing of laboratory supplies for sequencing runs.
- Presented honors research orally in several settings, including the North Eastern Association of Forensic Scientists and American Academy of Forensic Sciences.
- Collected, cataloged, and managed human samples and the required informed consent paperwork.

Culinary Leader
Redifer Dining Commons
The Pennsylvania State University, University Park, PA Spring 2015 – Spring 2018

- Improved positions starting as a student worker, to a culinary apprentice, and then to a culinary leader.
- Balanced a 20-hour work week while being a full time student.
- Assisted in management of multicultural student culinary team within a large working kitchen, while serving 1000 or more customers per meal.
- Created and maintained student culinary team schedules.
- Wrote and presented monthly meetings.
- Trained fellow students in team management, kitchen safety, knife handling, food preparation, and general cooking skills.

PROFFESSIONAL PRESENATIONS

Forensic Science Student Research Exchange
8th Annual Meeting, University Park, PA April 6 – 7, 2018

- Poster presentation of: *Investigation and comparison of mtDNA heteroplasmy rates across various haplogroups using massively parallel sequencing*.

American Academy of Forensic Sciences
70th Annual Meeting, Seattle, WA February 19 – 24, 2018

- Oral presentation of: *Investigation and comparison of mtDNA heteroplasmy rates across various haplogroups using massively parallel sequencing*.
- Attended full workshop of: *Domestic violence and child abuse deaths.*
- Attended scientific, poster, and general sessions.

North Eastern Association of Forensic Scientists
43rd Annual Meeting, Pocono Manor, PA November 7 – 10, 2017

- Oral presentation of: *Investigation and comparison of mtDNA heteroplasmy rates across various haplogroups using massively parallel sequencing*.
- Attended scientific, poster, general, and plenary sessions.

SKILLS

- Public speaking.
- Team management.
- Microsoft Office suite.
- Human sample collection.
- Communication of scientific concepts, including proposal and report writing.
- Massively Parallel Sequencing (MPS) using Illumina® MiSeq machine and Nextera XT library preparation and MPS data analysis using GeneMarker® HTS.
- Polymerase Chain Reaction (PCR) DNA amplification.
- DNA quantification using Qubit and Quantifiler® HP analyses.
- DNA extraction using organic, Chelex®, organic, differential, and Gentra® Puregene methods.
- Short Tandem Repeat (STR) DNA analysis using Applied Biosystems 3130xl Genetic Analyzer and PowerPlex® Fusion 6C kit.
- STR data analysis using GeneMarker® HID software.
- Phylogenetic analysis using GenBank published data and Molecular Evolutionary Genetics Analysis (MEGA) using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods
- Forensic, microbiology, cadaver, and biochemistry laboratory common practices.
- Presumptive and confirmatory testing of biological fluids.
- Analysis of hair and fibers.
- Processing of fingerprints and footwear impressions.
- Forensic photography and microscopy.

ACADEMIC HONORS AND AWARDS

Dean's List, The Pennsylvania State University Fall 2014 – Spring 2018
Schreyer Honors Scholar Fall 2014 – Present

EXTRA CURRICULAR EXPERIENCES

Knitvism President
Student Organization
The Pennsylvania State University, University Park, PA Fall 2016 – Spring 2018

- Spearheaded a creative organization that raises awareness of selected causes.
- Taught beginners the craft of crochet.
- Taught intermediate and advanced level crochet techniques.
- Coordinated donations to multiple non-profit organizations benefitting homeless, sexual assault victims, Lyme disease awareness, hospitalized children, and others.
- Coordinated with other student organizations to plan charitable events.
- Maintained Facebook group, Facebook page, and sites.psu.edu/knitivism during tenure.
- Held the offices of Secretary for two semesters, Vice-President for one semester, and President for four semesters.
- As secretary, maintained email listserv and wrote weekly meeting updates.