THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

EBERLY COLLEGE OF SCIENCE FORENSIC SCIENCE PROGRAM

EXAMINING MITOCHONDRIAL DNA HETEROPLAMSY IN HUMAN HAIR SHAFTS USING 454 LIFE SCIENCES NEXT GENERATION SEQUENCING

KERRY MCGINLEY
Spring 2012

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:

Dr. Mitchell Holland
Director, Forensic Science
Thesis Supervisor
Honors Adviser

Dr. Ronald Porter
Director, Graduate Studies
Faculty Reader

Abstract

Through the use of high resolution pyrosequencing the mitochondrial DNA sequence of hair shafts were analyzed for the existence of heteroplasmy. The primer set regions used in forensic mtDNA testing, two and three, were analyzed for five individuals using the 454 GS Junior instrument. Previous studies in our laboratory have shown the reliability of the GS Junior to detect heteroplasmy. This current study focuses on one of the most popular samples, hair shafts that forensic laboratories perform mtDNA testing on. The hair shafts also show the highest rate of heteroplasmy among all other common forensic sample types. The primer set regions were amplified using polymerase chain reaction (PCR) and the primers used to amplify the mtDNA had multiplex identifiers (MIDs) for sample identification and adapter sequences for the pyrosequencing process. After sequencing, the sequence data were analyzed using NextGENe software. The Next Generation 'primer set 3' primers used with the MID sequences did not amplify the hair shaft extract DNA due to the unforeseen existence of stable secondary structures of the primers. The Next Generation 'primer set 2' primers successfully amplified and sequenced the hair shaft extract DNA. However, there were problems with getting an appropriate coverage or number of reads to examine low-level heteroplasmy. Contamination was an initial concern because of the increased sensitivity of Next Generation sequencing, however results showed minimal contamination and the correct mtDNA profile was generated using the 454 GS Junior to perform the sequencing.

Table of Contents

Introduction

Human mitochondrial DNA has become an important tool in forensic DNA analysis and identification (Holland and Parsons 1999). There are advantages that mitochondrial DNA has over nuclear DNA that makes mtDNA identification testing possible with degraded samples, such as hair shafts, where nuclear DNA is limited. Mitochondrial DNA is maternally inherited (Giles et al 1980) which helps make identification of human remains possible when only distant relatives are available for comparison (Holland et al 1993; Sullivan et al 1992). There are up to thousands of copies of mtDNA within each cell (Bodenhagen et al 1974; Michaels et al 1982) which makes mtDNA testing useful if samples have gone through severe degradation or environmental insult. While there are only two copies of nuclear DNA per nucleated cell, the increased copy number of mtDNA makes it more likely to find useful mtDNA for testing. Finally, within the mtDNA sequence is a non-coding, hypervariable region called the control region which contains more polymorphisms than the coding region of mtDNA that can be used to distinguish one maternal lineage from another (Wilson et al 1993).

A matter that complicates yet strengthens analysis of mtDNA is the existence of heteroplasmy. Heteroplasmy is the occurrence of two or more sequences of mtDNA within a mitochondrion or cell, creating the occurrence of multiple mtDNA sequences within an individual. One of the first reports of heteroplasmy was in 1983 in a study examining non-coding mtDNA from human placentas (Greenberg 1983). The first recognition of heteroplasmy in forensic mtDNA testing was of skeletons of the Romanov family which were found in a shallow grave in Russia (Gill et al 1994; Ivanov et al 1996). While in general, the mtDNA sequence within an individual is consistent with the sequence inherited from the mother, there are mechanisms that cause sequences to vary. The low fidelity of the mtDNA polymerase causes mutations in an individual which are then amplified as mitochondria replicate (Kunkel and Loeb 1981). Reactive oxygen species generated from oxidation-reduction reactions that take place in mitochondria also damage mtDNA, which leads to mutations. The two hypervariable regions in the control region (the non-coding region) of the mtDNA genome are the areas where the majority of heteroplasmy is found because it is an area that is more susceptible to mutations as it does not code for genes (Stoneking 2000). Hypervariable region I (HVI) extends from base pair 16024 to about 16365 and hypervariable region II (HVII) extends from around base pair 73 to

340 (Holland and Parsons 1999). The existence of heteroplasmy can help differentiate individual members within the same family (Wilson et al 1993; Parsons et al 1997) if there is a difference in the levels of heteroplasmy among family members (Gocke et al 1997; Bendall et al 1996). Mitochondrial DNA evidence can be more discriminating when heteroplasmic sequences between samples match.

There has been much research done that shows that heteroplasmic ratios differ within an individual's body. Older studies have shown that there exist different ratios of heteroplasmy between different tissues within an individual caused by a genetic bottle neck (Hauswirth and Laipis 1982). This bottle neck creates variability among family members (Bendall et al 1997). A recent study showed that heteroplasmy is partially caused by a bottleneck, but most of the nucleotide differences are caused by somatic mutations during embryonic development (He et al 2010). Many studies have been done that suggest that individual hair strands have significant variance in heteroplasmic ratios (varying degrees of the more prevalent mtDNA sequence) among forensically relevant samples. Mechanisms that create heteroplasmy in hair may be more complicated than in other tissues and may begin with segregation of mtDNA types during regular tissue differentiation of the embryo and then go through a secondary bottle neck during development of the hair follicle (Melton 2004). The migration of mitochondria along the hair shaft causes complications in interpreting mtDNA between different samples from the same individual because you might have different heteroplasmic ratios of mtDNA along the same hair shaft or between hair shafts of an individual (Linch et al 2001). For example, an individual might have two mtDNA sequences throughout their hair shafts, a sequence with a C more prevalent and at a specific nucleotide and another sequence with a T more prevalent at the same nucleotide. If the individual leaves behind a hair at a scene with a high number of mtDNA sequences with the C nucleotide the profile could be reported as inconclusive if their hair or buccal reference had a high number of mtDNA sequences with the T nucleotide. There have been studies done that observed heteroplasmy in both hair and blood with higher rates of heteroplasmy in hair (Paneto et al 2007; Sekiguchi et al 2004). A study by Melton et al (2005) was done on 691 casework hair samples where heteroplasmy was observed 11.4% f the time.

Sanger dideoxynucleotide sequencing has been the main technique in sequencing mtDNA in forensic casework. In this method all of the mtDNA sequences in a sample are sequenced and

averaged together, making it difficult to deconvolute mixtures of different mtDNA sequence. The current Next Generation Sequencing technology allows for significantly increased sensitivity of detection for heteroplasmy compared to Sanger sequencing. The 454 Life Sciences platform utilizes a pyrosequencing approach. This technology allows each individual mtDNA template in a sample to be sequenced separately, like a large cloning experiment (Mardis 2008). Then each DNA molecule that was sequenced can be looked at and analyzed individually in order to detect heteroplasmy at lower frequencies. There have already been studies done to examine the increased discriminating power and sensitivity in detecting heteroplasmy in mtDNA using Next Generation Sequencing (Tang et al 2010 and Zaragoza et al 2010).

The 454 NGS approach to pyrosequencing involves two amplification reactions, purification, and sequencing by flowing nucleotides sequentially across a PicoTiterPlate device holding thousands of beads with DNA product. The second amplification is called emulsion PCR (emPCR) where individual DNA strands are attached to single beads which are then emulsified in the reaction reagents in a water-in-oil mixture. The single DNA molecule on the bead is amplified and the products are captured on the same bead, creating a clonally amplified population of the single DNA molecule. All the beads enriched with DNA from emPCR are then added to the PicoTiterPlate that has about 100,000 reaction wells large enough to fit one bead. As with all pyrosequencing technologies, four nucleotides are then flowed sequentially across the plate and when a base is added to a complementary nucleotide a coupled reaction with sulfurylase and luciferase release light that is captured by a CCD camera. The sequences can then be looked at individually in a software program and analyzed in order to detect low level heteroplasmy. The pyrosequencing approach gives greater sensitivity in detecting heteroplasmy in mtDNA. This technology provides enhanced detection and resolution of heteroplasmic variants in samples, allowing for the detection of low-level heteroplasmy, uncovering even more heteroplasmy in individuals than what was previously possible.

The purpose of the experiments in this research project centered on the concept that using the 454 GS Junior NGS platform to perform mtDNA sequence analysis would allow for greater detection of mtDNA heteroplasmy in human hair shafts. Preliminary studies done in our laboratory showed the ability of 454 NGS platform to sequence mtDNA, deconvolute mixtures, and allow a closer look at the content and rate of low-level heteroplasmy (Holland et al 2011).

Reproducibility studies have shown that sequencing is reproducible when detecting low-level heteroplasmy variants. The studies have also established thresholds and criteria for the reporting of low-level heteroplasmy variants to distinguish true heteroplasmy from PCR or sequencing errors. The importance of detecting low-level heteroplasmy is great in the forensic DNA community since there are limitations on the discrimination power of mtDNA testing using the standard Sanger sequencing. With the possibility to tap into low-level heteroplasmy through Next Generation sequencing mtDNA analysis of HVI and HVII could provide better discrimination from individual to individual in the future, possibility even between those of the same maternal lineage. Given that hair is a forensically relevant and important sample for mtDNA testing, the results of this work will be of great interest to the forensic science community. Research has shown that hair has the greatest occurrence of heteroplasmy out of all other forensically relevant samples. The problem is that Sanger sequencing limits our detection of low-level heteroplasmy. With the use of the 454 NGS platform and its ability to detect heteroplasmy, we hope to increase the discrimination potential of the testing system, even between different individuals in the same maternal lineage.

Methods

Genomic DNA was extracted from hairs by washing the hairs in an ultra-sonicator for three-20 minute washes in a terg-a-zyme solution and then in ethanol and water. The hairs were incubated at 56 degrees Celsius (ºC) with Proteinase K and 1 M DTT for two hours, and then the samples were incubated at 99 ºC for 10 minutes to inactivate the Proteinase K. The DNA was then isolated and purified with a Microcon 100 with salmon sperm coating. Fresh 1 M DTT was made each time extractions were completed. Routine amplification and Sanger sequencing was completed on all hairs extracted to get the mtDNA sequence for the ‘primer set 2’ and ‘primer set 3’ locations from each of the samples.

Common among forensic laboratories that are doing mtDNA testing, there are different amplification and sequencing strategies depending on type of sample. For samples with a high quantity of high quality DNA, each of the two hypervariable regions can be amplified in two independent PCR reactions. For samples that have lower quantities of possibly degraded DNA,

each of the two hypervariable regions are amplified in two overlapping regions for a total of four independent PCR reactions. The primers that are used to amplify the four regions across HVI and HVII are designated as 'primer set 1', 'primer set 2', 'primer set 3', and 'primer set 4'. Hair shaft samples are amplified using these four primer set regions. The current study focused on the 'primer set 2' location in HVI (ranging from nucleotides 16159 to 16401) and attempted to look at 'primer set 3' location in HVII (ranging from nucleotides 29 to 285) because these regions are home to the greatest number of mutational hotspots in the hypervariable region. The primer sequences designed for this study include primer binding sequences around the primer set regions, a multiplex identifier (MID) sequence that is used to differentiate samples being run on the same sequencing plate, and a fusion primer sequence that is used to bind the amplicons to the individual beads in the second round of amplification and to prime for sequencing. The Next Generation primers designed to include the MID and fusion primer sequencing for 'primer set 2' and 'primer set 3' are shown in Table 1. Five samples were run on a sequencing plate at once, so five different MIDs were used. The PCR reaction included 5 uL of hair extract, 10X PCR buffer, 10 mM dNTPs, 10 pM of each primer, 1 uL of bovine serum albumin (BSA), and 2.5 units of Taq Gold. The cycling parameters included an 11 minute, 95 ºC soak and then 38 cycles of 94 ºC for 1 minute, 60 ºC for 1 minute, 72 ºC for 1 minute. Twenty percent of the PCR reaction volume was then electrophoresed through a 2% agarose gel at 120 V to get an estimate of the quantity of amplicon. The PCR products were then purified using Amicon Ultra- 0.5 (with a 10 kDa cutoff) centrifugal filters to get rid of primers and PCR reagents before the second round of amplification. The entire PCR product along with 400 uL of water were added to the filter and spun in a centrifuge at 2,400 rotations per minute (RPMs) for 12 minutes. After the first spin, another 400 uL of water was added to the filter and spun again at 2,400 RPM for 12 minutes. The RPMs were eventually increased to 4,000 RPMs for 12 minutes for both spins to decrease the volume of the purified PCR product.

The samples were then diluted to equal concentrations and pooled together using the following approach. Each sample was diluted to $1X10^9$ molecules by an equation found in the GS Junior Titanium Amplicon Library Preparation Method Manual. Next, 10 uL of each product were pooled together to create a 50 uL volume aliquot. This pool was then diluted by a factor of 100 by combining 2 uL of the product pool and 198 uL of distilled water. Finally, a 10-fold

dilution was performed by combining 3 uL of the diluted product pool and 27 uL of distilled water for a total 1,000-fold dilution of the pooled DNA.

The five samples that were pooled together were then amplified by emulsion PCR (emPCR) by using the GS Junior Titanium emPCR (Lib-A) kit. During this process, the forward and reverse products from each of the five samples are attached to beads, with one single stranded DNA molecule per bead. Through the emulsion process the single molecule captured to a single bead is amplified, creating multiple copies of the same mtDNA molecule per bead. After emPCR was completed, the beads were then captured, purified and enriched. The emulsion was broken with a vacuum and then the beads were washed and recovered. During the enrichment process, beads with mtDNA molecules attached to them were collected and sorted out from the beads that did not contain DNA. A GS Junior emPCR bead counter device was used to estimate the number of enriched beads. An output of 500,000 to 2 million beads is optimal for GS Junior pyrosequencing.

The enriched beads were then prepared for sequencing using the GS Junior Titanium Sequencing Kit. The beads were added to the PicoTiterPlate along with sequencing reagents, and DNA polymerase. The plate, along with the appropriate buffers, was added to the GS Junior Instrument. After the sequencing, the data were analyzed on Next GENe software from Soft Genetics.

Five hair shaft samples were extracted for mtDNA and run on the GS Junior. The first two experiments failed to give enough coverage to examine low-level heteroplasmy. It was determined that the GS Junior properly sequenced the hair shaft samples and gave the same mtDNA profile when compared to Sanger sequencing. When dealing with low levels of DNA extracted from hair shaft samples, there is a greater possibility of contamination in the samples. With the increased sensitivity of Next Generation sequencing there was a concern about picking up a lot of contamination that Sanger sequencing would not detect. There is currently a run that is being conducted to examine low-level heteroplasmy with a sufficient amount of coverage.

Table 1. The primer sequences for the first round of polymerase chain reaction (frPCR) for 'primer set 2' and 'primer set 3'. The red indicates the fusion primer sequence, the green indicates the MID sequence, and the blue indicates the primer binding sequence.

Next Generation 'Primer Set 2' Primer Sequence

MID 1 Forward	5'- GCC TCC CTC GCG CCA TCA G ACG AGT GCG T TAC TTG ACC ACC TGT AGT AC – 3'
MID 1 Reverse	3' –GT GGT AGG AGG CAC TTT AGT TGC GTG AGC A GAC TCG CCC GAC CGT TCC G – 5'
MID 9 Forward	5'- GCC TCC CTC GCG CCA TCA G TAG TAT CAG C TAC TTG ACC ACC TGT AGT AC – 3'
MID 9 Reverse	3' - GT GGT AGG AGG CAC TTT AGT CGA CTA TGA T GAC TCG CCC GAC CGT TCC G – 5'
MID 10 Forward	5'- GCC TCC CTC GCG CCA TCA G TCT CTA TGC G TAC TTG ACC ACC TGT AGT AC – 3'
MID 10 Reverse	3' - GT GGT AGG AGG CAC TTT AGT GCG TAT CTC T GAC TCG CCC GAC CGT TCC G – 5'
MID 11 Forward	5'- GCC TCC CTC GCG CCA TCA G TGA TAC GTC T TAC TTG ACC ACC TGT AGT AC – 3'
MID 11 Reverse	3' - GT GGT AGG AGG CAC TTT AGT TCT GCA TAG T GAC TCG CCC GAC CGT TCC G – 5'
MID 12 Forward	5'- GCC TCC CTC GCG CCA TCA G TAC TGA GCT A TAC TTG ACC ACC TGT AGT AC – 3'
MID 12 Reverse	3' - GT GGT AGG AGG CAC TTT AGT ATC GAG TCA T GAC TCG CCC GAC CGT TCC G – 5'

Next Generation 'Primer Set 3' Primer Sequence

MID 1 Forward	5'- GCC TCC CTC GCG CCA TCA G ACG AGT GCG T GT CTA TCA CCC TAT TAA CCA C – 3'
MID 1 Reverse	3' – G TTT TTT AAA GGT GGT TTG GGG TGC GTG AGC A GAC TCG CCC GAC CGT TCC G – 5'
MID 9 Forward	5'- GCC TCC CTC GCG CCA TCA G TAG TAT CAG C GT CTA TCA CCC TAT TAA CCA C – 3'
MID 9 Reverse	3' - G TTT TTT AAA GGT GGT TTG GGG CGA CTA TGA T GAC TCG CCC GAC CGT TCC G – 5'
MID 10 Forward	5'- GCC TCC CTC GCG CCA TCA G TCT CTA TGC G GT CTA TCA CCC TAT TAA CCA C – 3'
MID 10 Reverse	3' - G TTT TTT AAA GGT GGT TTG GGG GCG TAT CTC T GAC TCG CCC GAC CGT TCC G – 5'
MID 11 Forward	5'- GCC TCC CTC GCG CCA TCA G TGA TAC GTC T GT CTA TCA CCC TAT TAA CCA C – 3'
MID 11 Reverse	3' - G TTT TTT AAA GGT GGT TTG GGG TCT GCA TAG T GAC TCG CCC GAC CGT TCC G – 5'
MID 12 Forward	5'- GCC TCC CTC GCG CCA TCA G TAC TGA GCT A GT CTA TCA CCC TAT TAA CCA C – 3'
MID 12 Reverse	3' - G TTT TTT AAA GGT GGT TTG GGG ATC GAG TCA T GAC TCG CCC GAC CGT TCC G – 5'

Results

Five hair shaft samples were extracted and amplified and sequenced using Sanger technique with 'primer set 2' and 'primer set 3' primers that did not include the MID or fusion primer sequence. The Sanger mtDNA profile of the five samples is shown in Table 2. In the past there has been trouble with hair shaft extractions in our laboratory. Mainly the DTT was performing inadequately to reduce the disulfide bonds in the hair shafts. This led to the inability to break down the hair shafts. It was important to be sure that the hair shaft extractions were working properly before proceeding with GS Junior sequencing.

Table 2. The mtDNA profile of five hair shaft samples for 'primer set 2' and 'primer set 3'regions.

Samples	**Sanger mtDNA profile (primer sets 2 and 3)**	**454 GS Junior mtDNA profile (primer set 2)**
M1	16189 C 16193.1 C 16265 C 16291 T	16189 C 16265 C 16291 T
M2	73 G 185 A 228 G	No polymorphisms
F1	16224 C 16311 C 73 G 146 C 152 C	16224 C 16311 C
F2	73 G	No polymorphisms
F3	16193 T 16263 C 16294 T	16263 C 16294 T

Once it was certain that the hair shaft extractions were working properly the hair shaft samples were amplified with Next Generation primers for both 'primer set 2' and 'primer set 3' regions. While the five forward and reverse primers designed for the 'primer set 2' region properly amplified the mtDNA in the hair shaft extracts, as shown in the agarose gel results, the 'primer set 3' region primers did not amplify the mtDNA properly, as shown in negative gel results.

The Next Generation ‘primer set 3’ primers were designed and worked with first. The problems with these primers lead to the development of a set of PCR conditions that amplified mtDNA from the hair shaft samples, using the Next Generation ‘primer set 2’ primers, successfully. The first attempt at amplifying the hair shafts with the Next Generation ‘primer set 3’ used conditions designated in our laboratory’s amplification of mtDNA protocol. The total reaction volume was 50 uL with 1 uL each of the forward and reverse primer and 5 uL of the mtDNA extract added to the reaction. The hair shaft samples were never quantified for the amount of nuclear or mtDNA within the sample. The cycling parameters included a 12 minute, 96ºC soak and then 38 cycles of 95 ºC for 15 seconds, 56 ºC for 30 seconds, 72 ºC for 45 seconds. The agarose gel showed no bands for each of the five PCR products. After there were no results from the above conditions a second amplification was run where the total reaction volume was decreased and the amount of DNA extract and the primer added to the reaction were increased. The total reaction volume was 25 uL with 2.5 uL each of the forward and reverse primers. This reflects the reagent and reaction volumes used previously in our laboratory for pristine buccal samples that underwent sequencing with the GS Junior. A positive control was also run under the MID 12 primers of ‘primer set 3’. The same thermal cycle parameters for the first amplification were used as well. The agarose gel again showed no bands for the five amplified hair shaft samples as well as the positive control.

In order to determine what in the PCR reaction was not working, the same conditions as the second amplification were kept constant, except for the concentration of the dNTPS. The concentration used in all previous reactions was diluted 100-fold which could have been a limiting reagent in the reaction. A new concentration of 10 mM dNTPs were used in the amplification reaction. Again, the agarose gel showed no results. After determining that all the reagents used in the PCR reactions were working properly through other reactions preformed, it was determined that the Next Generation ‘primer set 3’ primers were not working properly in the PCR conditions and parameters that were previously used.

After the first troubleshooting process was over and the problem with the amplification of the hair shafts was identified as the Next Generation ‘primer set 3’ primers, a second troubleshooting process began. Different experimental conditions in the PCR reaction itself and the cycling conditions on the thermal cycler were changed in order to get results from the Next

Generation ‘primer set 3’ primers. All experiments were performed using pristine buccal swab samples because the chance of the primers amplifying an extract with pristine and a large amount of DNA as opposed to a smaller quantity of DNA extracted from a hair shaft would be greater. The amplification thermal cycling parameters that were used previously in our laboratory for buccal samples that underwent sequencing with the GS Junior were used. The previous studies used primers of similar length (around 50 base pairs) and of a similar melting temperature of around 60 ºC. The new cycling parameters included longer denaturing, annealing and extension times as well as a higher annealing temperature. The new cycling parameters for the amplification of the mtDNA hair shafts included an 11 minute, 95 ºC soak and then 38 cycles of 94 ºC for 1 minute, 60 ºC for 1 minute, and 72 ºC for 1 minute. In the fourth amplification reaction the Next Generation ‘primer set 3’ primers were used to amplify buccal swab samples that have amplified in the past. The samples contained an average of 13 ng/uL of nuclear DNA. The cycling parameters described above were used with the 25 uL total reaction volumes with 1 uL of extract added. The agarose gel showed no results once again.

The final modification to the PCR thermal cycling conditions in an attempt for the Next Generation ‘primer set 3’ primers to successfully amplify the region of the mtDNA was to increase the annealing temperature due to the high melting temperature of the primers. The melting temperature of the five forward and reverse primers ranged between 69.3 ºC and 72.2 ºC and averaged 70.3 ºC according to Integrated DNA Technologies, the company the primers were ordered from. The annealing temperature was raised to closer to the melting temperature of the primers, 68 ºC. The same thermal cycling parameters, reagent volumes, and reaction volume as the fourth amplification reaction were used. A total of 19.06 ng of DNA from a buccal swab were added to the reaction to be amplified. With the increase in annealing temperature the agarose gel still yielded no results.

After consistently getting no results by amplifying with the Next Generation ‘primer set 3’ primers, some primer artifact was detected after adding dimethyl sulfoxide (DMSO) to the PCR reaction to get rid of potential secondary structure that the primers might be experiencing. To the 25 uL total reaction, 10% pure DMSO, 2.5 uL was added to the reaction. A pristine buccal sample was amplified with all the primers and the DMSO. The agarose gel showed faint bands around the 100 base pair region for MID 1, MID 10, and MID 11 primers, shown in Figure

1. Because the bands were around the 100 base pair region, the product is most likely the result of primer dimer formation and not actual product, which would be expected to be around the 350 base pair region. As each primer is around 50 base pairs in length, a primer dimer situation could show bands at around the 100 base pair region. All of the amplification reactions performed are summarized in Table 3.

Figure 1. Primer artifacts shown through the existence of bands detected on an agarose product gel around 100 base pairs in length for MID 1,MID 10, and MID 11. In the MID 9 and MID 12 lanes, no product was detected.

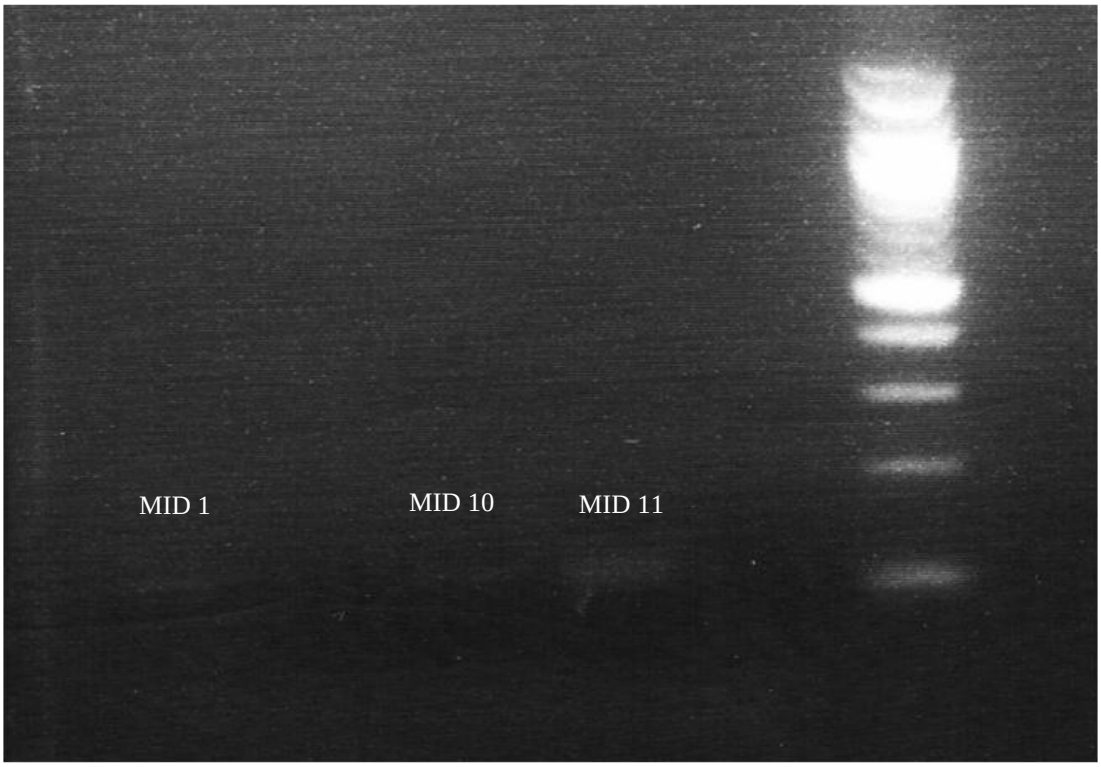


Table 3. Summary of amplification reactions performed including the type of sample added, the total volume of DNA extract added, the amount of forward and reverse primers added, the total reaction volume of the PCR reaction, the thermal cycling parameters used, and the gel results.

Amplification Reaction	Total DNA Extract Added (uL)	Amount of uM Primer Added (uL)	Total Reaction Volume (uL)	Cycling Parameters	Gel Results
1	5	1	50	• 12 minute 96ºC soak • 38 cycles • 95 ºC for 15 seconds • 56 ºC for 30 seconds • 72 ºC for 45 seconds	No bands detected
2	5	2.5	25	• 12 minute 96ºC soak • 38 cycles • 95 ºC for 15 seconds • 56 ºC for 30 seconds • 72 ºC for 45 seconds	No bands detected
3	10	2.5	25 (new concentration of dNTPs added)	• 12 minute 96ºC soak • 38 cycles • 95 ºC for 15 seconds • 56 ºC for 30 seconds • 72 ºC for 45 seconds	No bands detected
4	1 uL Samples included 7.22 ng/uL, 9.53 ng/uL, 20.85 ng/uL of nuclear DNA for an average of 13 ng/uL	2.5	25	• 11 minute 95ºC soak • 38 cycles • 95 ºC for 1 minute • 60 ºC for 1 minute • 72 ºC for 1 minute	No bands detected
5	2 uL 9.53 ng/uL of nuclear DNA	2.5	25	• 11 minute 95ºC soak • 38 cycles • 95 ºC for 1 minute • 68 ºC for 1 minute • 72 ºC for 1 minute	No bands detected
6	2 uL 9.53 ng/uL of nuclear DNA	2.5	25 (2.5 uL of DMSO added to the reaction)	• 11 minute 95ºC soak • 38 cycles • 95 ºC for 1 minute • 60 ºC for 1 minute • 72 ºC for 1 minute	MID 1, MID 10, MID 11showed bands for primer artifacts

After adding the DMSO to the PCR reaction it was determined that the primers were experiencing secondary structure such as hairpins and dimers and therefore would not anneal to the mtDNA to amplify the ‘primer set 3’ regions in the extract. The DMSO did help three sets of primers to break the hairpin structure; however, the primer dimer formation still did not allow the

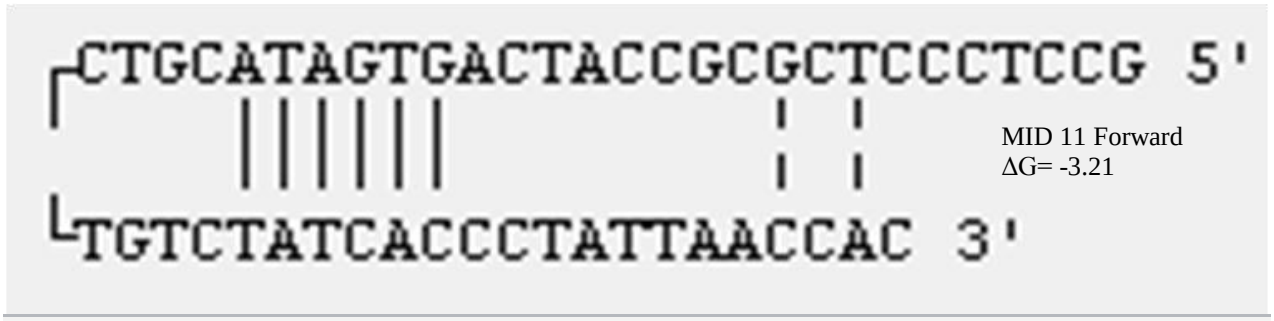
primers to properly anneal to the template DNA. Table 4 shows the properties of the Next Generation 'primer set 3' primers as well as their secondary structure properties. The rating makes it easy to predict the efficiency of a primer and takes into the primary types of secondary structure, hairpins and dimers. The higher the rating number, the more efficient the primer will be. For each individual primer, rating = 100 + (ΔG(dimer) X 1.8 + ΔG(hairpin) X 1.4).The guanine and cytosine content (GC content) is determined by dividing the sum of the guanine and cytosine nucleotides in the primer by the total number of bases in the primer and then multiplying by 100 to get a percent. The 3' end ΔG and the 5' end ΔG is calculated by determining the free energy (ΔG) of the last five bases on the 3' end and the 5' end of the primer, respectively. The number of hairpins and the number of dimers include the number of potential secondary structures that can occur with each primer. Figure 2 shows the hairpin and dimer structures with the lowest free energy, or the secondary structures with the greatest potential in decreasing the amplification efficiency. The free energy of the hairpins and dimers is calculated to determine the stability of the secondary structure formed. The more negative the free energy is, the higher the potential is to remain as a secondary structure and effect the efficiency of amplification.

Table 4. Next Generation 'primer set 3' primers properties.

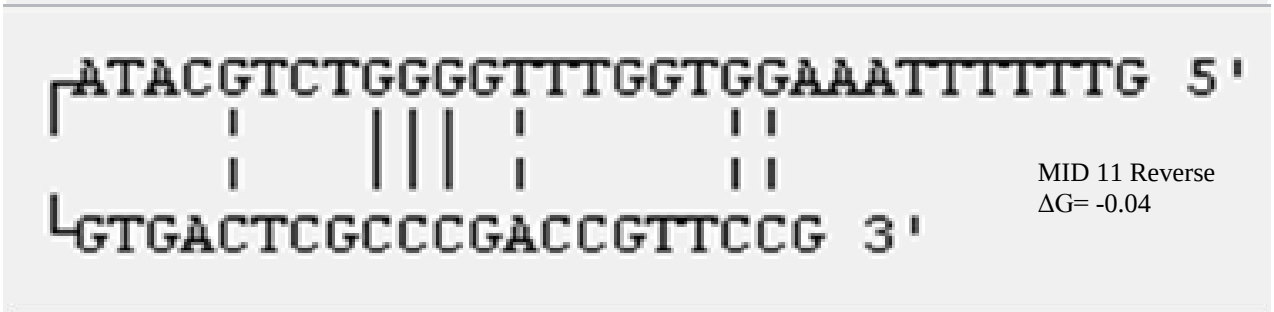
Primer	**Rating**	**Melting Temp (ºC) at 50 mM NaCl**	**GC%**	**3' end ΔG (kcal/mol)**	**5' end ΔG (kcal/mol)**	**Number of Hairpins**	**ΔG of Hairpins (kcal/mol)**	**Total Dimers (3' dimers in parentheses)**	**ΔG of Dimers (kcal/mol)**
MID 1 Forward	77	71.4	58.0	-7.71	-9.38	5	-2.48	7 (2)	-10.36
MID 1 Reverse	80	72.2	56.86	-10.2	-7.18	3	-2.45	1	-8.74
MID 9 Forward	79	69.9	54.0	-7.71	-9.38	2	-1.15	3	-10.36
MID 9 Reverse	59	69.9	52.94	-10.2	-7.18	1	-10.16	2	-14.46
MID 10 Forward	79	70.4	56.0	-7.71	-9.38	3	-1.55	4 (1)	-10.36
MID 10 Reverse	65	71.0	54.9	-10.2	-7.18	2	-8.39	2	-12.89
MID 11 Forward	76	69.5	54.0	-7.71	-9.38	1	-3.21	7 (1)	-10.36
MID 11 Reverse	84	70.4	52.94	-10.2	-7.18	1	-0.04	3	-8.74
MID 12 Forward	70	69.3	54	-7.71	-9.38	1	-3.53	6	-13.9
MID 12 Reverse	44	69.8	52.94	-10.2	-7.18	2	-8.13	5 (1)	-24.79

Figure 2. Hairpin and dimer structures for the five Next Generation ‘primers set 3’ primers, shown with the free energy (ΔG) calculations in kcal/mol.

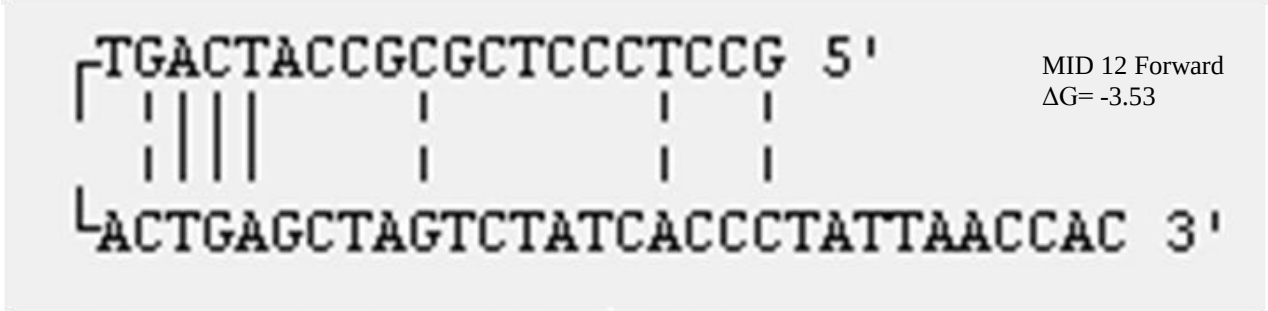
Hairpin Structures



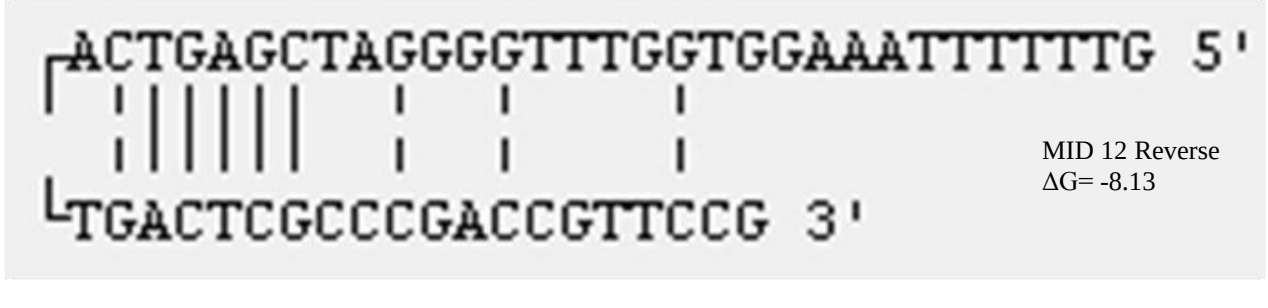
MID 11 Forward
ΔG= -3.21



MID 11 Reverse
ΔG= -0.04



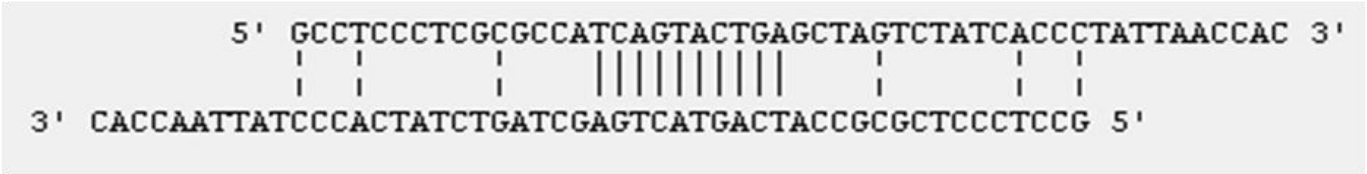
MID 12 Forward
ΔG= -3.53



MID 12 Reverse
ΔG= -8.13

Dimer Structures

Fusion primer and PS3 primer sequence creating the dimer in the MID 1, 9, 10, 11 forward primers (MID 11 shown)
ΔG= -10.36



MID 12 Forward
ΔG= -13.9

```
5' GTTTTTTAAAGGTGGTTTGGGGCGACTATGATGACTCGCCCGACCGTTCCG 3'
              ||     | |||||| |      | |||||| |     ||
              ||     | |||||| |      | |||||| |     ||
         3' GCCTTGCCAGCCCGCTCAGTAGTATCAGCGGGGTTTGGTGGAAATTTTTTG 5'
```

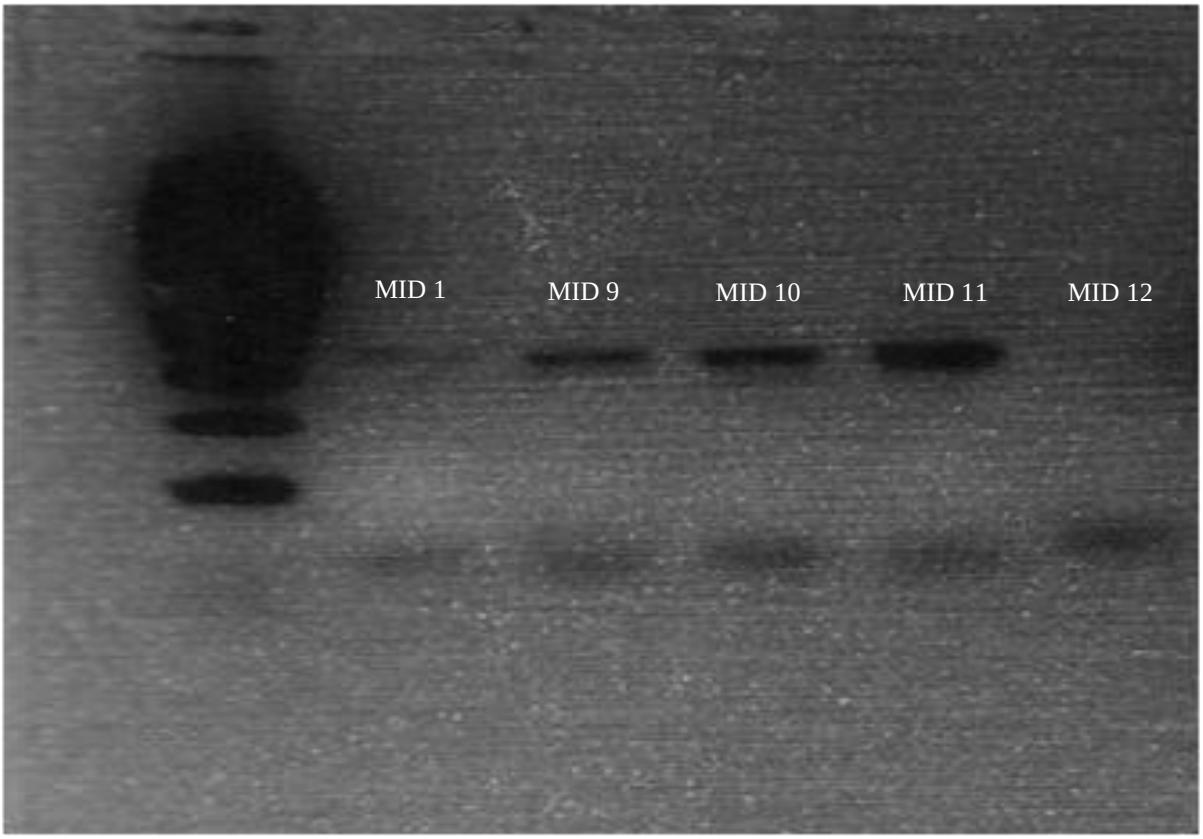
MID 9 Reverse
ΔG= -14.46

```
5' GTTTTTTAAAGGTGGTTTGGGGATCGAGTCATGACTCGCCCGACCGTTCCG 3'
              |     |  |  ||||||||||||||  |  |     |
              |     |  |  ||||||||||||||  |  |     |
         3' GCCTTGCCAGCCCGCTCAGTACTGAGCTAGGGGTTTGGTGGAAATTTTTTG 5'
```

MID 12 Reverse
ΔG= -24.79

The first round of amplification with the Next Generation 'primer set 2' primers yielded agarose gel results for four out of the five samples, as shown in Figure 3. Although in the figure shown there is no MID 12 result, sequences were obtained from this sample.

Figure 3. Product gel of hair shaft samples amplified with the Next Generation 'primer set 2' primers. The upper band is the product band and the lower band is from primer artifacts.



After purifying and diluting and pooling the samples from the five MIDs together as described in the Methods section, the enriched beads were estimated with the GS Junior emPCR bead counter device. The beads appeared to be slightly below the recommended 500,000 beads which are optimal for pyrosequencing. For this first run, sequencing still proceeded but data analysis showed low coverage and low reverse reads for the samples. The coverage ranged from 36 to 907 reads with an average of 521 reads across the five samples and a total coverage of 2,667. The average read length was adequate, 231 across the five samples. The low coverage was caused by too much dilution of the samples. After the amount of PCR product was determined by the band intensity on the agarose gel, the samples were purified to get rid of the primers. The purification process is thought to be the step that caused the samples to be too diluted by further diluting the product beyond the calculation in the Methods section; the final purification products were around 80 uL as opposed to the 25 uL amplification reaction volumes. During the purification for this first run, the samples were spun at 2,400 RPMs.

The first run was re-done with the same samples. The extracts were re-amplified for the first round of amplification. When the products were purified the samples were spun at 4,000 RPMs and the samples were not diluted by the calculations discussed in the Methods section. Instead, the products were individually diluted to get $1X10^9$ molecules, then 10 uL of each product were pooled together and the pool was only diluted by 100-fold as opposed to 1,000-fold that was done in the previous run. The enriched beads were estimated with the bead counter device and there were far more than two million beads as a result of not diluting the samples enough. The run could not proceed. The process had to be repeated once more, to get proper bead count between 500,000 and 2 million enriched beads.

The dilutions of the purified PCR products that were spun down at 4,000 RPMs were re-done according to calculations described in the methods section. This was in response to getting over 2 million beads by only diluting the pool 100-fold. After spinning the products by 4,000 RPMs, diluting each product to $1X10^9$ molecules, pooling 10 uL of each product together and then diluting the pool 1,000-fold, the proper amount of enriched beads were obtained, just over 500,000. Sequencing proceeded for a second run on the GS Junior.

The results of the first two runs are shown in Table 5. In the first run one of the samples was discarded as Sanger sequencing results were not obtained. The total coverage was low for

both runs, even though the second run had the optimal number of enriched beads. The first run had a total coverage of 2,667 while the second run had a total coverage of 2,856. The 'primer set 2' regions cover around 242 base pairs in length, so the average read length of 231 for the first run and 224 for the second run was around what would be expected. When comparing the Sanger sequencing mtDNA profile to the Next Generation results, the same profile for each sample was obtained. There is concern about sequence reliability around homopolymeric stretches as the pyrosequencing approach tends to add or delete nucleotides around these regions. Due to this concern, the insertions in the M1 DNA profile, 16193.1 C, could not be reported reliably. Overall, the positions of 16166 and 16183 showed nucleotide A deletions consistently across all the profiles. Positions 16188, 16192, and 16193 also had a large number of insertions and deletions because it is homopolymeric region. Any true polymorphism or low-level variants are lost within these regions.

When looking at the profiles generated from sequencing with the GS Junior, contamination is not a major issue when reporting the mtDNA profile with most of the samples observed. There was a concern of picking up more contaminants because the technology is more sensitive. This problem is exacerbated when dealing with DNA from hair shafts, as there is a smaller amount of DNA present in hair shafts. In the first run, in all cases, the majority of nucleotides at a specific position gave the correct profile, as shown in Table 5. Because of the low coverage and the low number of reverse reads, the sequences showing the minority percent of nucleotides at the polymorphic positions (the last column in Table 4) cannot be determined as low-level heteroplasmy or sequencing error. In general, out of all the possible sequencing errors across the profiles, insertions and deletions of nucleotides were the most common. For the first run with sample M1, it appears that out of the 36 sequences, two of them were contaminants from either M2 or F2. These two sequences, which matched the reference sequence, were the reason for the minority of other nucleotides at the positions of polymorphisms (see Table 5.). In all of the samples except for M1.1 and F3, the minor nucleotide variants caused by heteroplasmy or sequencing errors were rarely more than 5% of the total type of nucleotide at a specific base M1.1 and F3 both appear to be a mixture or have high levels of contamination shown by the nucleotide ratios for the positions of polymorphisms. The nucleotide position of 16193 was not considered for sample F3 because of the high amount of sequencing artifacts (insertions and deletions of nucleotides) around that region due to homopolyermic stretches.

Table 5. Results of the two sequencing runs on the GS Junior from the hair shaft samples. The total coverage for each base includes the number of DNA molecules sequenced. The average read length shows the average number of nucleotides sequenced amongst all the DNA molecules. The last two columns show the percent of nucleotides shown at that position location along the sequence. The 'Nucleotide % of Polymorphism' column shows the percentage of the particular nucleotide polymorphism at that particular position amongst all the DNA molecules. The 'Nucleotide % of possible low-level variant or sequencing errors' shows the percentage of other nucleotides sequenced at that particular position amongst all the DNA molecules. The top table shows results for the first run and the bottom table shows results for the second run.

Sample	**Total Coverage**	**Average Read Length**	**mtDNA profile for primer set 2**	**Nucleotide % of Polymorphism**	**Nucleotide % of possible low-level variant, sequencing errors, or contamination**
M1	36	232	16189 C 16193.1 C 16265 C 16291 T	16189 C- 94.44 16265 C- 91.69 16291 T- 94.44	16189 T- 5.56 16265 A- 5.56 16291 C- 5.56
M2	907	237	No polymorphism	N/A	N/A
F1	866	235	16224 C 16311 C	16224 C- 98.73 16311 C- 99.88	16224 T- 1.04, A- 0.12 16311 T- 0.12
F2	653	238	No polymorphism	N/A	N/A
Sample	**Total Coverage**	**Average Read Length**	**mtDNA profile for primer set 2**	**Base % of Polymorphism**	**Base % of possible low-level variant or sequencing errors, or contamination**
M1.1	2011	200	16189 C 16193.1 C 16265 C 16291 T	16189 C- 45.28 16265 C- 61.99 16291 T- 62.12	16189 T- 39.56 16265 A- 37.34, G and T- 0.05 16291 C- 37.65, A and G- 0.05
M2.1	285	216	No polymorphisms	N/A	N/A
F1.1	228	229	16224 C 16311 C	16224 C- 100 16311 C- 98.23	16311 T- 1.77
F2.1	252	236	No polymorphisms	N/A	N/A
F3	80	238	16193 T 16263 C 16294 T	16193 T- 50.62 16263 C- 85.19 16294 T- 86.42	16193 C- 49.38 16263 T- 1.23 16294 T- 13.58

Discussion

When designing the primers for Next Generation sequencing for 'primer set 2' and 'primer set 3', there was no initial concern of secondary structure. The primer binding sequences used were the same as the sequences used for 'primer set 2' and 'primer set 3' for Sanger sequencing which have always worked in our laboratory in the past, and we have not experienced this problem when designing next generation primer in past experiments. The primers were ordered from Integrated DNA Technologies where their secondary structure calculations showed no indication that secondary structure should affect the yield or purity for each of the primers synthesized. Prior work in our laboratory focused on HVI primers (Holland et at 2011). Therefore, considerable time was spent trouble shooting and determining the cause of the problems with 'primer set 3'.

When amplifying the hair shaft extracts with the Next Generation 'primer set 3' primers, the reaction and reagent volumes and cycling parameters used the first time around were the same that our laboratory uses when sequencing mtDNA with the Sanger technique. The PCR reaction volume was 50 uL and only 1 uL of the primer was added to the reaction. When no product was shown in the agarose gel, the total reaction volume was cut in half to 25 uL. With the decrease in reaction volume, which made the extract more concentrated, and the increase in the amount of primer added to 2.5 uL the possibility of results increased. For the second amplification the amount of extract 20% of the total volume as opposed to 10% that is usually used for Sanger sequencing was added, so plenty of extract was being added. By increasing the amount of primer it made sure it was not a limiting reagent in the reaction.

After it was determined that all of the other PCR reaction components were working properly the focus was on allowing the primers to properly anneal to the template DNA and amplify. Previous amplification reactions with Next Generation primers that amplified the whole HVI region (which included the fusion primer and MID sequence) used a higher annealing temperature of 60 ºC and a longer annealing time, 1 minute as opposed to 30 seconds. The annealing temperature was increased because it was closer to the melting temperature of the Next Generation 'primer set 3' primers that averaged around 70.3 ºC. The annealing temperature was increased because the Next Generation primers are long. They are around 50 base pairs in length as opposed to other primers for Sanger sequencing which are around 20 base pairs in length. The

thought was that increasing the annealing time would give the primers more time to bind to the template DNA, which was important because of the length. After no results were seen, the annealing temperature was increased by 8 ºC to 68 ºC to be even closer to the melting temperature of the primers. The idea was this would help them anneal better by preventing non-specific binding of the primers, and reducing secondary structure of the individual primers. After no results were obtained, the secondary structure of the primers were examined more closely.

Dimethyl sulfoxide is used in PCR reactions to interfere with the self-complementarily of the PCR primers to inhibit secondary structure before they bind to the template DNA and to enhance specificity by speeding the binding of the primer to the template DNA. The DMSO helps to disrupt base pairing to some degree. After the previous experiments were run and the troubleshooting focused on potential secondary structure of the primers, pure DMSO was added to the PCR reaction at 10% of the total reaction volume; in this case 2.5 uL for the 25 uL reactions. Although the agarose gel results were mostly weak, adding DMSO to the reaction helped to get primer product for three out of the five Next Generation ‘ primer set 3’ primers: MIDs 1, 10, and 11 to break the hairpin structures of the primers. The MID 9 and 12 primers have hairpin and dimer structures with the most negative free energy values (ΔG); they have the most stable hairpin and dimer structures. A hairpin is formed when the primer folds back on itself and relies on complementary regions along the primer and is stabilized by intramolecular bonding. A self- dimer is formed when there is homology present within a primer and is formed by two copies of the same primer binding to each other.

The free energy of the secondary structures of a primer determines the stability of the hairpin or dimer formed by the primer. The free energy is the energy required to break the secondary structure, with a larger negative value signifying the more stable the secondary structure is. The overall rating of the primer, shown in Table 4, shows the efficiency of the primer. A rating closer to 100 is a more efficient primer. The rating number takes the free energy of the worst potential hairpin and dimer secondary structures in consideration, with the free energy of the dimer weighed slightly more than the free energy of the hairpin. The MID 9 and 12 reverse primers have the worst ratings of all the primers with a rating of 59 and 44, respectively. These primers have extremely stable dimer formation shown by the free energy of their reverse primers, -14.46 kcal/mol for MID 9 and -24.79 kcal/mol for MID 12. For the MID 9

primer, shown in Figure 2, there are two occasions of six bases in a row finding complementary bases in the primer to create a dimer. The MID 12 has even more stable dimer formation with an occasion of fourteen bases in a row finding complementary bases in the primer. It seems that the DMSO could not help interfere with the stable secondary structures of the MID 9 and 12 primers.

For the forward Next Generation 'primer set 3' primers, the dimers for MIDs 1, 9, 10, and 11 are formed from four complimentary bases in a row between the fusion primer sequence and the primer binding region. The dimer is not terribly stable, with a free energy of -10.36 kcal/mol. For the MID 12 forward primer, the dimer formed was more stable, with a free energy of -13.9 kcal/mol formed from eleven complimentary bases in a row due to the MID sequence.

Included in the rating of primer efficiency is the free energy of the potential hairpins formed by the primers. The MID 9 and MID 12 Next Generation 'primers set 3' primers, reverse primers, had some of the most stable hairpin structures compared to the other primers. The MID 9 reverse primer had six complementary bases in a row, while most of the primers had hairpins formed by three or four complementary bases in a row, shown in Figure 2. The free energy of -10.16 kcal/mol was the most stable hairpin structure formed out of all the primers. For the MID 12 reverse primer, the free energy of -8.13 kcal/mol was the third most stable hairpin structure formed out of all the primers; six complementary bases in a row caused the formation of the hairpin structure. Through the stable self-dimer and hairpin formation, the MID 9 and 12 primers could not break free secondary structure with the addition of 10% DMSO. The MID 1, 10, and 12 primers have some potential still of binding to the template DNA. There are some other reagents that could be added to help break the secondary structure further. Betaine might be a useful reagent to potentially add to the PCR reaction to help break the dimer formation as it is used to reduce secondary structure, especially in GC rich regions. All of the primers looked at have realitively high GC content. Formamide could potentially help the primers bind to the template DNA. It is has also been shown to increase the PCR efficiency on difficult secondary structures of primers.

The melting temperatures of the primers, although high, are relatively similar. The annealing temperature of 60 ºC worked for the Next Generation 'primer set 2' primers which were of similar melting temperature. So, it seems that melting temperature does not affect the

efficiency of the Next generation ‘primer set 3’ primers. The GC% content of the primers averages to 54.7%, within the 40 to 60 % range of acceptable primers, although on the higher end.

The 3’ end ΔG it determines the primer’s false priming efficiency. The more stable the 3’ end is, the greater the chance it will bind to another sequence rather than its intended target. The lower the free energy of the 3’ end of the primer, the greater the chance the primer will bind to non-specific sites. In the Next Generation ‘primer set 3’ primers, the forward primers all had a free energy of -7.71 kcal/mol and the reverse primers all had a free energy of -10.2 kcal/mol. Since the 3’ end of the primer is the primer binding sequence taken from the Sanger ‘primer set 3’ primers, the free energy would be the same. The Sanger primers have always worked in amplification and sequencing so the 3’ end ΔG, although lower, should not have an effect on the efficiency of the primers. The free energy of the (5’) end determines the stability of the primer binding to the complementarily region on the target DNA. The more stable the 5’ end is, the more efficient the primer is because it ensures adequate binding of the primer to the template DNA. The five prime ends of the primers are the fusion primer sequence; the same sequence has been used in our laboratory in past primers, so again that should not affect the efficiency of the Next Generation ‘primer set 3’ primers

While the Next Generation ‘primer set 2’ primers amplified the hair shaft mtDNA template DNA well, there were problems with getting the appropriate number of enriched beads because of the dilutions of the PCR product. In past runs on the GS Junior in our laboratory, the PCR product was never purified to get rid of primer and PCR artifacts before the emPCR process. The PCR product was purified because, in the agarose gel, results showed some primer dimer formation by way of a secondary band other than the PCR product, farther down the gel, shown in Figure 3. The PCR products were purified at first by spinning the samples with 400 uL of water at 2,400 RPMs which led to a diluted purified PCR product with a final volume of around 80 uL. The dilution calculations described in the methods section did not consider further dilution by the purification process, leading to a product that was too diluted; in this instance the number of enriched beads was lower than the optimal 500,000 beads for sequencing. By increasing the RPMs of the centrifuge during the purification process, the final volume of all the purified PCR products were around 35 uL, meaning that the products were less diluted and closer

to the concentration of the 25 uL PCR reaction volumes. After increasing the RPMs during the purification and following the dilution calculations described in the methods section, the appropriate number of enriched beads for sequencing was obtained. Spinning the PCR products at 2,400 RPMs and then diluting the purified products by the calculations described in the methods section led to products that were too diluted and a low number of enriched beads. By increasing the RPMs to 4,000 and only diluting the pooled products by 100-fold, the products were not diluted enough and led to a high number of enriched beads, too many for sequencing. By getting the purified PCR product volume closer to the 25 uL reaction volume, which would only slightly dilute the products and following the calculation described in the methods section the appropriate number of beads were obtained.

In the two experiments, the coverage or amount of DNA molecules sequenced were both too low to detect low-level heteroplasmy variants. From previous studies in our laboratory when performing runs on the GS Junior, a recommended 40 reads must be generated with a consistent ratio of forward to reverse reads compared to the total ratio in order to reliably report low-level variants. In order to reliably report low-level variants down to 1%, then, a coverage of at least 4,000 reads must be obtained. At the very least, the results showed that sequencing hair shaft samples can show reliable results through sequencing with the 454 NGS platform, although insertions and deletions in the sequences have to be looked at more carefully because of the high amounts of insertion and deletion sequencing errors. Because the intensity of the light is proportionate to the number of nucleotides added, homopolymeric stretches are difficult to sequence using pyrosequencing because the CCD camera has trouble detecting the intensity after about four of the same nucleotides are added in a row, resulting in insertions and deletions of nucleotides.

It was expected that the first run would have lower coverage because of the low amount of enriched beads. However, the second run yielded low coverage of 2,856 even though the optimal number of enriched beads was added to the sequencing reaction, which would anticipate 70,000 reads. Experiments are currently being conducted to try to increase the number of reads for the DNA molecules extracted from the hair shafts. So, while our results have shown that the 454 NGS platform can reliably sequence hair shaft samples, more studies need to be done to increase the coverage in order to tap into true low-level variants, and the source of low-level

contamination must be addressed. In order to reliably assess the difference between potential contaminants and heteroplasmy, this will be a critical piece of information as this project moves forward.

References

1. Holland MM and Parsons TJ. Mitochondrial DNA Sequence Analysis- Validation and Use for Forensic Casework. For Sci Review 11(1); 1999.

2. Giles RE, Blanc H, Cann HE, Wallace DC. Maternal Inheritance of human mitochondrial DNA. Proc. Natl. Acad. Sct. 77(11): 6715-6719; 1980.

3. Holland MM, Fisher DL, Mitchell LG, Rodriquez WC, Canik JJ, Merril CR, Weedn VW. Mitochondrial DNA sequence analysis of human skeletal remains: Identification of Remains from Vietnam War. J Forensic Sci 38 (3): 542-553; 1993.

4. Sullivan KM, Hopgood R, Gill P. Identification of human remains by amplification and automated sequencing of mitochondrial DNA. Int J Leg Med 105: 83-86; 1992.

5. Bogenhagen D, Clayton DA. The Number of Mitochondrial Deoxyribonucleic Acid Genomes in Mouse L and Human HeLa Cells. J Biol Chem 249:7991-7995; 1974.

6. Michaels GS, Hauswirth WW, Laipis PJ. Mitochondrial DNA copy number in bovine oocytes and somatic cells. Dev Biol 94(1):246-251;1982.

7. Wilson MR, Stoneking M, Holland MM, DiZinno JA, Budowle B. Guidelines for the use of mitochondrial DNA sequencing in forensic science. Crime Lab Digest 20(4):68-77; 1993.

8. Greenberg BD, Newbold JE, Sugino A. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene 21(1-2): 33-49; 1983.

9. Gill P, Ivanov PL, 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 6: 130-135; 1994.

10. Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. Nature Genetics 12: 417-420; 1996.

11. Kunkel TA, Loeb LA. Fidelity of mammalian DNA polymerases. Science 213 (4509): 765-767; 1981.

12. Stoneking M. Hypervariable Sites in the mtDNA Control Region Are Mutational Hotspots. Am J Hum Genet 67: 1029-1032; 2000.

13. Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, Berry DL, Holland KA, Weedn VW, Gill P, Holland MM. A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15: 363-368; 1997.

14. Gocke CD, Benko FA, Rogan PK. Transmission of mitochondrial DNA heteroplasmy in normal pedigrees**.** Hum Genet 102 (2):182-186; 1997.

15. Bendall KE, Macaulay VA, Baker JR, Sykes BC. Heteroplasmic point mutations in the human mtDNA control region. Am J Hum Genet 59(6):1276-1287; 1996.

16. Hauswirth WW, Laipis PJ. Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. PNAS 79(15):4689-4690; 1982.

17. Bendall KE, Macaulay VA, Sykes BC. Variable levels of a heteroplasmic point mutation in individual hair roots. Am J Hum Genet 61(6): 1303-1308; 1997.

18. He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, Diaz Jr LA, Kinzler KW, Vogelstein B, Papadopoulos N. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells; Nature 464: 610-614; 2010.

19. Melton T. Mitochondrial DNA Heteroplasmy. Forensic Sci Rev 16(1); 2004.

20. Linch CA, Whiting DA, Holland MM. Human Hair Histogenesis for the Mitochondrial DNA Forensic Scientist. J Forensic Sci 46(4): 844-853; 2001.

21. Paneto GG, Martins, JA, Longo, LVG, Pereira GA, Freschi A, Alvarenga VLS, Chen B, Oliveria, RN, Hirata MH, Cicarelli RMB. Heteroplasmy in hair: Differences among hair and blood from the same individuals are still a matter of debate. For Sci Inter 173(2):117-121; 2007.

22. Sekiguchi K, Sato H, Kasai K. Mitochondrial DNA Heteroplasmy Among Hairs from Single Individuals. J Forensic Sci 49(5); 2004.

23. Melton T, Dimick G, Higgins B, Lindstrom L, Nelson K. Forensics Mitochondrial DNA Analysis of 691 Casework Hairs. J Forensic Sci 50(1): 1-8; 2005.

24. Mardis ER. Next-Generation DNA Sequencing Methods. Annu. Rev. Genomics Hum. Genet. 9:387-402; 2008.

25. Tang S, Huang T. Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system. BioTechniques 48:287-296; 2010.

26. Zaragoza MV, Fass J, Diegoli M, Lin D, Arbustini E. Mitochondrial DNA Variant Discovery and Evaluation in Human Cardiomyopathies through Next-Generation Sequencing. PLoS ONE 5(8): e12295; 2010.

27. Holland MM, McQuillan MR, O'Hanlon KA. Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy. Croat Med J 52: 299-313; 2011.

Kerry McGinley

5812 Village Lane, Doylestown, PA 18902
kmcginley613@gmail.com
(215) 290-7970

EDUCATION

Pennsylvania State University University Park, PA

Master of Professional Studies in Forensic Science: Biology Expected May 2012
Bachelor of Science in Forensic Science: Biology with a Biochemistry/Molecular Biology minor

Related Completed Coursework: Crime Scene Investigation, Criminalistics: Trace Evidence and Serology, Molecular Biology of the Gene, Introduction to the Criminal Justice System, Biochemistry, Cellular and Molecular Biology, Genetics, Biostatistics, Laboratory in Proteins, Nucleic Acids, and Molecular Cloning, Forensic Molecular Biology, Courtroom Testimony, Molecular and Cellular Toxicology, Ethics in Forensic Science, Forensic Toxicology

Research: Examining mitochondrial DNA heteroplasmy in human hair shafts using 454 Life Sciences Next Generation Sequencing

Member of the Schreyer Honors College, Pennsylvania State University

WORK EXPERIENCE

Teaching Assistant, Pennsylvania State University, Forensic Science Program, World Campus
Online Forensic Crime Scene Investigation – September 2010 – Present

> Mentoring students through online meetings and email, discussing crime scene investigation techniques (crime scene management, photography, trace evidence, blood spatter interpretation, fingerprint evidence, evidence collection, biological fluid analysis, bullet trajectory analysis), grading assignments

Teaching Assistant, Pennsylvania State University, Forensic Science Program, University Park
Forensic Crime Scene Investigation – August 2009 – May 2010

> Teaching workshops regarding crime scene investigation techniques (fingerprinting, proper use of a camera, wet and dry residue prints, collection of evidence, blood spatter analysis, sketching), setting up crime scenes

Research Assistant, Pennsylvania State University, Biology Department, University Park – January 2009 – May 2009

> Polymerase chain reaction and running gels of DNA extracts of herpetological species, sequencing DNA, maintaining the laboratory

ADDITIONAL SKILLS

Processing and developing fingerprints/footwear impressions
Forensic Photography
Forensic Microscopy
Blood Spatter Analysis
Presumptive and Confirmatory testing of biological fluids
Analysis of hair and fibers
DNA Extraction
    Chelex, Organic, Differential, AutoMate Express Forensic DNA Extraction System and PrepFiler Forensic DNA Extraction Kit from Applied Biosystems, hair shafts
Quantification
    Applied Biosystems 7500 Real-Time PCR System
    Applied Biosystems Quantifiler Human and Quantifiler Y Human Male DNA Kits
Amplification of STR and Y STR Loci
    Applied Biosystems 9700
    Applied Biosystems Identifiler and Yfiler PCR Amplification Kits
    Applied Biosystems Identifiler Plus PCR Amplification Kit
Separation and Detection of Amplicons
    Applied Biosystems 3130xl Genetic Analyzer
Analysis of Data
    GeneMapper ID-X v1.3 from Applied Biosystems
    GeneMarker HID from SoftGenetics
    Mutation Surveyor from SoftGenetics
Forensic DNA Statistics
    Random Match Probability
    Combined Probability of Exclusion/Combined Probability of Inclusion
    Y-STR and mtDNA Statistics
Proper Case File Management
    Photo and written documentation
    Forensic DNA report writing
Next Generation Sequencing
    454 Life Sciences GS Junior System
    NextGENe software by SoftGenetics (State College, PA)

ACTIVITES AND ACHIEVEMENTS

Selected as Student Marshal of the Penn State Forensic Science Program	Spring 2012
Treasurer of the Penn State Forensics Club	August 2009- 2012
Awarded Wieland Scholarship	Summer 2010
Awarded Holtzinger Scholarship in Science	Summer 2010
Member of Rules/Regulations Committee for Penn State's Dance MaraTHON	2008-2009
Member of Penn State's Science Lions	2008-2009
Awarded President's Freshman Award, Pennsylvania State Abington Campus	Winter 2007