Koenig, Jessica. Evaluation of Molecular Techniques Using a Synthetic Mitochondrial Genome
Master of Science (Biomedical Sciences, Forensic Genetics). May, 2014.

84 pp. 16 tables, 29 figures, references

The mitochondrion is responsible for the bulk of cellular energy production through the process of oxidative phosphorylation. The mitochondrial genome (mtGenome) is subject to a high mutation rate due to its proximity to reactive oxygen species produced in energy production. Over 250 pathogenic mutations have been characterized, and studies have demonstrated mtDNA variations at the cellular, tissue, and individual level. Some of the characterization techniques include long range PCR and sequencing. Sanger sequencing has been the gold standard, but next-generation sequencing technologies are now available. These methods may be evaluated using synthetic DNA of known base composition. This project utilizes the first synthetic mtGenome to optimize a LR PCR protocol and evaluate sequence quality using Sanger, MiSeq System, and Ion Personal Genome Machine System sequencing platforms.

Evaluation of Molecular Techniques

Using a Synthetic

Mitochondrial Genome

Jessica Koenig, B.S.

APPROVED:

_____

Rhonda Roby, PhD, MPH, Major Professor

_____

Michael Allen, PhD, Committee Member

_____

Robert Barber, PhD, Committee Member

_____

Dong-Ming Su, PhD, University Member

_____

Arthur Eisenberg, PhD, Chair, Department of Molecular and Medical Genetics

_____

Meharvan Singh, PhD, Dean, Graduate School of Biomedical Sciences

EVALUATION OF MOLECULAR TECHNIQUES USING A SYNTHETIC MITOCHONDRIAL GENOME

THESIS

Presented to the Graduate Council of the

Graduate School of Biomedical Sciences

University of North Texas

Health Science Center at Fort Worth

Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Jessica Koenig, B.S.

Fort Worth, Texas

May 2014

ACKNOWLEDGEMENTS

I would first like to thank my wonderful husband who uprooted his life and supported me both emotionally and financially in the pursuit of my dream. To my major professor, Dr. Rhonda Roby, I am extremely grateful. I requested a project that would allow me to "get my hands dirty" in the lab, and she met that request in spades. Her dedication and attention to detail elevated the quality of my thesis project to a level I would not have accomplished on my own. I give my sincerest thanks to Nicole Phillips and Marc Sprouse for their invaluable feedback and assistance. I thank my committee members, Dr. Michael Allen, Dr. Robert Barber, and Dr. Dong-Ming Su at the University of North Texas Health Science Center for their guidance throughout this project. I would also like to thank Dr. Daniel Gibson and his colleagues at the J. Craig Venter Institute whose synthesis of the mitochondrial genome was the inspiration for this project. Finally, I offer thanks to all my family and friends for their love and support.

Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

The mitochondrion is the cellular powerhouse, and is responsible for the bulk of a cell's energy production through the process of oxidative phosphorylation (OXPHOS) (1). Critical to this process is the mitochondrial genome (mtGenome). The mtGenome was sequenced in 1981, and was the first genome completely sequenced (2). A revised version determined in 1999 is still utilized today and is referred to as the revised Cambridge reference sequence (rCRS) (3). The approximately 16,569 base pair (bp) circular genome encodes 37 genes and structural RNAs, 13 of which encode proteins involved in the electron transport chain. It also contains a region of approximately 1125bp known as the 'control region' or 'D-loop'.

The control region contains regulatory elements for transcription and translation but does not code for proteins or structural RNAs and is therefore subject to fewer functional constraints. It harbors two hypervariable regions (HV1 and HV2) commonly utilized in forensic testing for missing persons cases, kinship analyses, and cases in which degradation renders nuclear DNA difficult or impossible to type. Sequence variations in the control region define an individual's mitochondrial haplotype that can be used to aid in the determination of identity and/or maternal relationships through comparison to known reference samples and geographic origin through databased haplotypes.

Sequence variation within the mtGenomes of an individual is referred to as heteroplasmy and may be characterized as differences in length or presence of point mutations. Single cells may contain anywhere from 200 to 1700 copies of the mtGenome (4) which are in close proximity to damaging free radicals produced during OXPHOS and bear no protective histones

(5). The mtGenome replicates independently with a higher error rate and less efficient repair than nuclear DNA. This results in a mutation rate approximately ten to 20 times that of nuclear DNA (6-9). Initial studies demonstrated little intra-individual variation in the mtGenome (10), and therefore, individuals were assumed to be basically homoplasmic. However, a recent study found heteroplasmic sites throughout the mtGenome present at varying levels between tissues of non-diseased individuals. This study shows that an individual does not have a single mitochondrial genotype and that the genotype of a single cell cannot be assumed to be homoplasmic (11). Therefore, it is possible that individuals have many differences in their mtGenome which may be present at varying levels, and detection of these differences is dependent upon a sensitive and accurate typing system.

Mitochondrial DNA mutations are associated with a variety of metabolic and neurological disease states. The first pathogenic mitochondrial DNA (mtDNA) mutations were identified in 1988 (12, 13), and over 250 pathogenic mutations are currently known (9, 14). The ability to distinguish mutations from errors produced by detection methodology is important in a forensic and clinical setting. In most cases, mutations are present in some but not all mtGenomes, and disease onset is determined by the ratio of mutant to wild type mtDNA. This threshold typically exists within a range of 60% to 90% mutant dominance, but can vary greatly by mutation type and the primary tissue affected. Large scale deletions may vary in size and have been reported to include up to 11kb (14). Clinically, deletions have historically been detected by Southern blotting, though long-range PCR (LR PCR) has replaced this technique as the method of choice as it is less labor-intensive and has a low cost (9). This technique is similar to traditional PCR amplifications with slight modifications to aid in the production of large amplicons. The LR PCR technique is used for amplicons as large as 30kb in length. The human

mtGenome is approximately 16.6kb and may be amplified in near entirety using LR PCR depending on where primer sets are located.

Analysis of LR PCR amplifications can be complicated by the existence of background PCR artifacts. This is demonstrated in a study of mitochondrial deletions in cardiomyocytes of middle-aged and elderly human donors in which the whole mtGenome was amplified from individual cells using the TaKaRa LA PCR system (TaKaRa Bio USA, Madison, WI) (15) (Figure 1). Background bands were seen to be of higher intensity than bands presumed to represent low level deletion products.

Background bands were also detected in preliminary experiments conducted at the UNTHSC. LR PCR was utilized for whole genome amplification of mtDNA not expected to contain deletions from three healthy donors. DNA was extracted from whole blood using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA) and amplified using the SequelPrep™ Long PCR kit (Life Technologies, Carlsbad, CA). Samples were amplified in triplicate. Intact mtDNA, the dominant 16kb product, was successfully amplified for each sample. Background bands were seen that were inconsistent between replicates of the same sample indicating that these products do not represent true deletions (Figure 2), and made it impossible to distinguish background bands from low level deletions based on visual examination of gel electrophoresis alone.

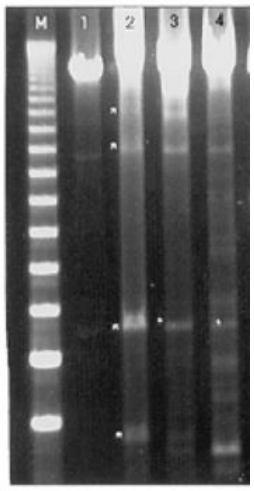


Figure 1. LR PCR of the MtGenome from Individual Cardiomyocytes.

M is a one kb size marker. Lane 1, a typical amplification from a cell that does not contain deleted mtDNA but displays presence of background bands. Lane 2, the same PCR product as lane 1 but overloaded to more clearly display background bands which are presumed to be PCR artifacts (marked by white asterisk). Lanes 3 and 4, PCR from DNA isolated from homogenized heart tissue of a 31 and a 101 year-old donor, respectively. Wells were overloaded to visualize low intensity bands. It is unclear if additional bands visible in lanes 3 and 4 are low level deletion products or additional PCR artifacts. Figure modified from Khrapko *et al* (15).

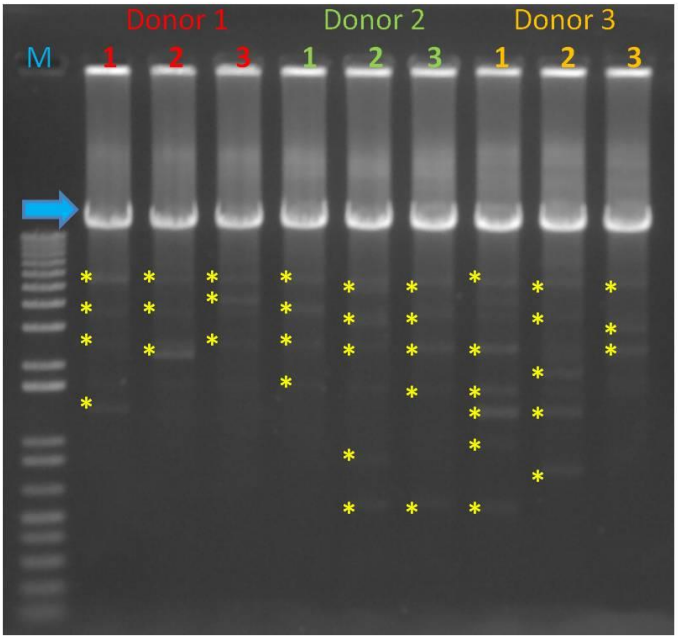


Figure 2. LR PCR at UNTHSC.

MtGenome amplification of whole blood from three healthy donors amplified in triplicate (Donor 1 in lanes 1-3 (red), Donor 2 lanes 4-6 (green), and Donor 3 lanes 7-9 (orange)). M is a 1 kb Plus DNA Ladder (Life Technologies). Intact mtDNA was successfully amplified for each sample, and is visible as the dominant product (aligned to the blue arrow). Faint background bands are also present in each amplification. Background bands are not consistent between replicate amplifications of the same sample (yellow asterisks).

The most common forensic analysis of mtDNA is sequence determination of the mtDNA haplotype through sequencing of the D-Loop which contains hypervariable (HV) regions, HV1 and HV2. This region is sequenced in the forward and reverse directions, evaluated for concordance, and the haplotype is reported as variants via comparison to the revised Cambridge Reference Sequence (rCRS) (16). In this regard, Sanger sequencing methods have been the gold standard for over thirty years (17). This method utilizes dideoxy chain termination coupled with fluorescent fluorophores (18). Products are separated via capillary electrophoresis during which

fluorophores are excited with an argon laser and detected with a charge coupled device (CCD) camera (19-22). Data interpretation software utilizes a *phred* algorithm which assigns base calls and quality scores that designate error probability. Scores are reported as the log-transformed probability that the base call is incorrect and is multiplied by -10. For example, a base call with a 1 in 1000 probability of error is assigned a quality score of Q30 (23, 24). Sanger sequencing is capable of producing read lengths up to 900 base pairs (Table 1) (25-28). Drawbacks to this method include low throughput and high costs. Also, both HV1 and HV2 contain poly-cytosine stretches that are poorly replicated and may result in multiple length heteroplasmies (16). When this occurs, it is difficult to sequence beyond this region because sequences are shifted out of synchronization by this variable homopolymeric region. In addition, this method has limitations in its ability to detect heteroplasmy. Heteroplasmy present at less than 20% is not easily distinguished (29). In instances of heteroplasmy present at lower levels, the minor component may be erroneously disregarded as sequence error or background noise. It is presumed that variants may be present at levels too minimal to be distinguished from background noise; as a result, either the predominant base is called and the minor peak is disregarded, or the nucleotide position is deemed inconclusive.

Table 1. Comparison of Sequencing Methodologies.

This information has been compiled from Liu *et al.* (30), Quail *et al.* (31) and Tonge *et al.* (32)

Sequencer	Mechanism	Average Read Length	Average Quality	Data Output
Capillary Electrophoresis	Dideoxy chain termination	400-900bp	Q20	1.9 to 84kb
MiSeq System	Sequencing by synthesis	300bp	Q30+	1.5 to 2GB
Ion PGM™ System	Semiconductor sequencing	400bp	Q20	20Mb to 1Gb

One study of 5,015 samples utilizing Sanger sequencing of the control region determined that the approach could detect heteroplasmy to a minor component level of approximately 10%. A total of 52% of the samples contained length heteroplasmy with the majority occurring in the poly-cytosine region of HV2 (45%) and a smaller proportion (15%) exhibiting length heteroplasmy in the poly-cytosine region of HV1. Length heteroplasmy was also commonly observed in the dinucleotide repeat region of individuals whose mtGenome contained more than five repeat units. However, due to limitations of the method, these heteroplasmic sites could not be confirmed as true variants as opposed to artificially introduced stutter products. Point heteroplasmies were also observed in 6% of samples with the majority (97%) exhibiting a single heteroplasmic site and a small portion (3%) exhibiting two or more (33).

Within the last several years, next generation sequencing (NGS) technologies have revolutionized genomic studies through their ability to produce massive amounts of data in a relatively short amount of time and with ever-decreasing costs. For example, the MiSeq System (Illumina, San Diego, CA), released in 2011, is a benchtop sequencer that can produce approximately 2Gb of 300bp paired-end sequence reads in a run time of about 65 hours (Table 1). This platform utilizes Illumina's sequencing by synthesis technology (Figure 3). DNA is

amplified, fragmented, and ligated to adapters. The DNA is then denatured to single strands and attached to a flow cell via the adapters. Isothermal 'bridge' amplification is then used to form clusters of approximately one million clonally expanded copies of the originally bound fragment. In this process, the denatured strands are copied and the original template is removed by denaturation. Then, the 3' end of the newly copied strand anneals to a second surface-bound oligonucleotide forming a bridge. The complementary strand is resynthesized. Multiple cycles of this process are repeated to form the clusters which are large enough to generate a detectable signal during sequencing. Sequencing is conducted using fluorescently labeled reversibly terminated nucleotides. At each cycle, all four nucleotides are added simultaneously. Fluorophores are excited via a diode laser (530 and 660 nm), and the signal is captured by a CCD camera. Tris(2-carboxyethyl)phosphine is used to remove the fluorescent dye and regenerate the 3'-hydroxyl group of the incorporated nucleotide in preparation for the next cycle of nucleotide addition. An algorithm is used to interpret the fluorescent signal and to apply base calls and quality scores. (34, 35)

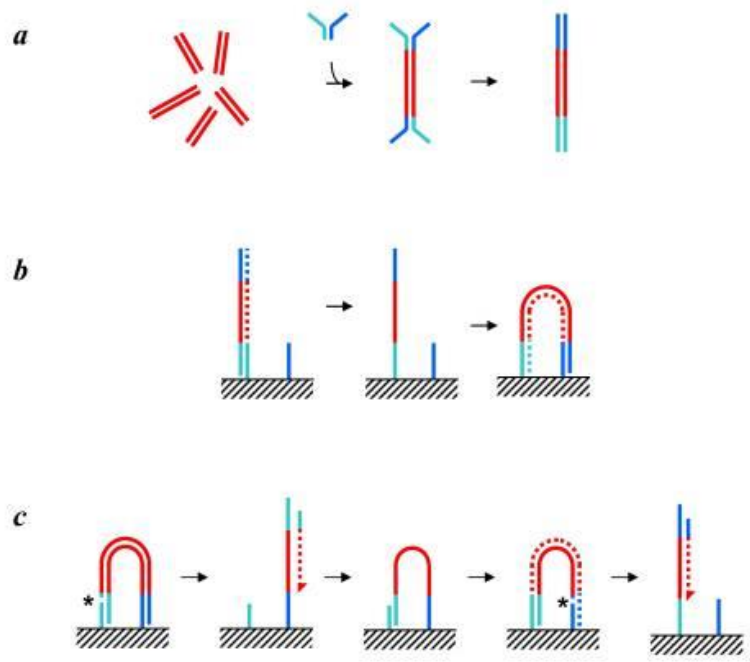


Figure 3. The MiSeq System Sequencing by Synthesis Technology.

a) DNA (red) is fragmented and ligated to adapters (blue) b) DNA is bound to the flowcell via adapters and denatured to single-strands followed by bridge amplification. In this process, the 3' end of the single-strand is bound to the flow cell. The surface oligonucleotide acts a primer, and the single-stranded template is copied (dotted line in the bridge). This process is repeated to form template clusters c) The template is linearized by cleavage at the adapter (asterisk) and denaturation of the complementary sequence. This provides a single-stranded template for sequencing by synthesis (dotted line). The synthesized DNA is removed by denaturation. The template again forms a bridge with a surface oligonucleotide. The complementary strand is resynthesized. The original template is cleaved at the adapter (asterisk) and removed by denaturation. Sequencing by synthesis is conducted on the complementary strand (dotted line) to generate the paired-end read. (Figure modified from Bentley et al.(34))

The Ion Personal Genome Machine (PGM™) System (Life Technologies) was also released in 2011, and is a platform that utilizes semiconductor sequencing technology (Figure 4). Sample DNA is first fragmented and ligated to adapters. Individual fragments are bound to beads. Template strands are clonally expanded via emulsion PCR. Then the beads are dispensed into the wells of a microchip. The chip is flooded with a single species of dNTP. The complementary dNTP is incorporated into the growing strand by polymerase and a hydrogen ion is released resulting in a pH change. An ion sensitive layer beneath the well converts the change in pH to a change in voltage which is reported to the software and the base call is assigned. The chip is flooded by one nucleotide after another. If the nucleotide is not incorporated, no voltage is reported. If two nucleotides are incorporated, the reported voltage will be doubled. Individual base calls are assigned quality values using an adapted *phred* model (36).

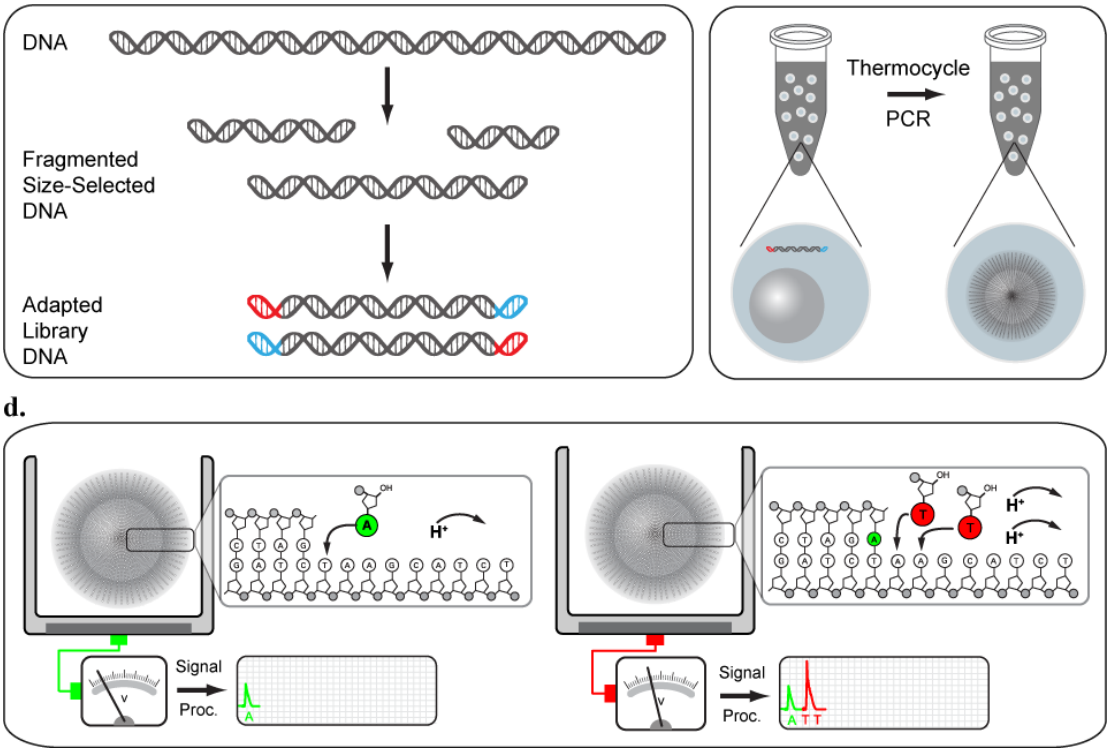


Figure 4. The Ion PGM™ System Semiconductor Sequencing Technology.

DNA is fragmented and ligated to adapters. Individual fragments are bound to beads and clonally expanded via emulsion PCR. The beads are dispensed into wells of the microchip. The chip is sequentially flooded by a single species of dNTP. Incorporation of a dNTP results in release of a proton causing a change in pH which is reported as a change in voltage. The pH change and resulting change in voltage is proportional to the number of dNTPs incorporated. (Figure modified from Rothberg *et al.* (36))

No sequencing method has shown to be one hundred percent accurate. Errors and sequence variations including insertions or deletions (indels) and mismatches are commonly identified via assessment of base quality and alignment to a reference sequence. Recent studies determine accuracy by evaluating concordance between multiple NGS platforms and/or Sanger results through reference mapping (31, 37, 38).

One study reported on the performance of the MiSeq System and Ion PGM™ System platforms by sequencing of a pathogenic *E. coli* and comparison to a reference genome produced by Roche 454 GS FLX+ System (Roche Diagnostics Corp., Indianapolis, IN) at an average depth

of 32-fold (38). The MiSeq System produced the highest quality reads with a rate of 0.1 substitutions per 100bp and less than 0.001 indels per 100bp. The Ion PGM™ System had 1.5 indels per 100bp, displayed decreased accuracy across the read, and showed poor performance in regions with homopolymeric stretches. The most commonly observed error in these regions was deletions, and this platform had as low as 60% accuracy for homopolymeric regions of 6bp or longer. Assemblies from both platforms were affected by unmapped regions, with 4.60% and 3.95% of the reference genome missing from Ion PGM™ System and MiSeq System, respectively.

A study by Quail *et al.* (31) also reported high accuracy for the MiSeq System, and observed an average base quality of Q30 and greater and a raw error rate of 0.80%. In contrast, the Ion PGM™ System produced an average base quality of Q20 and had a raw error rate of 1.71%. Overall, 76.45% of MiSeq System reads were error free, compared with 15.92% for the Ion PGM™ System. They also found significant bias when sequencing the AT-rich genome of *P. falciparum* using the Ion PGM™ System. As much as 30% of the genome had no coverage at all, and only 65% of the genome had high quality reads (>Q20); though, they did note that this was affected by the choice of polymerase used during amplification steps. The Ion PGM™ System failed to produce reads for homopolymeric stretch longer than 14bp, and could not accurately determine the number of bases in a homopolymeric stretch longer than 8bp. When sequencing this genome, errors were produced by the MiSeq System after long homopolymeric stretches greater than 20bp. Errors were noted for the GGC motif as well, particularly when associated with GC rich regions, but no error occurred if this motif was associated with an AT-rich region. In general, the MiSeq System produced data with fewer errors and was better able to sequence homopolymeric stretches than the Ion PGM™ System. Advantages to the Ion PGM™ System

include lower instrument costs and faster run times. As previously noted, the choice of polymerase enzyme utilized during library preparation and emulsion PCR can greatly affect accuracy and bias. Also, data interpretation can vary significantly among software programs and is dependent on subjective interpretation by the analyst. At present, formal quality assessment standards for analysis of NGS data are not in place (39).

In 2009, Dr. Daniel Gibson and colleagues at the J. Craig Venter Institute introduced a novel method for synthesizing and assembling DNA molecules which enables the assembly of synthetic DNA molecules up to several kilobases in size (40). This method, commonly referred to as the Gibson Assembly, utilizes the 5' T5 Exonuclease (Epicentre® Biotechnologies, Madison, WI), Phusion DNA Polymerase (New England Biolabs, Ipswich, MA), and *Taq* DNA Ligase (New England Biolabs) to allow for *in vitro* assemblage of overlapping oligonucleotides in a single isothermal step (Figure 5). The exonuclease 'chews' back the 3' end of a double-stranded molecule to create a single-stranded 5' overhang. This single-stranded overhang anneals to its complement on the overlapping oligonucleotide. The polymerase then fills in any gaps and nicks are sealed together by DNA ligase. The method may be utilized in a variety of molecular applications including cloning of PCR amplified fragments without restriction enzyme cut sites and site-directed mutagenesis.

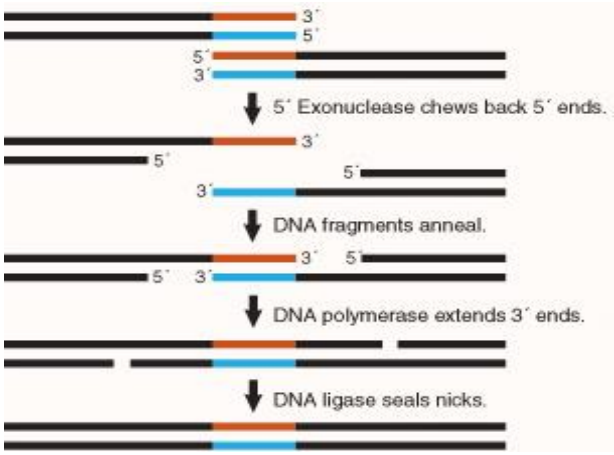


Figure 5. Overview of the Gibson Assembly.

The 5' ends of overlapping molecules are chewed by the exonuclease to generate 3' single-stranded overhangs. The overlapping overhangs anneal. Polymerase fills in gaps and nicks are sealed by DNA ligase. (Figure modified from New England Biolabs (41))

The method was developed as part of an effort to produce the first bacterial cell with a functioning synthetic genome. This was accomplished in 2010, when Dr. Gibson and colleagues synthesized and assembled the 1.08Mb *Mycoplasma mycoides* genome. The genome was successfully transplanted into *M. mycoides* cells. Cells containing only the synthesized genome were demonstrated to be self-replicating and exhibited characteristics of wild type *M. mycoides* cells (42). Also in 2010, Gibson and colleagues utilized this method to assemble the first synthetic mammalian organelle genome. They began with 600 overlapping 60-mer oligonucleotides as dictated by the sequence of the *Mus musculus* mtGenome (GenBank record NC_005089) and assembled the 16.3kb mouse mtGenome using four subassembly steps (Figure 6). The genome was assembled into a pUC19 cloning vector and used to transform *E. coli* cells. The genome is flanked on either side with a 221bp repeat that can be utilized for re-circularization without the pUC19 vector following its removal by the *Pml*I restriction enzyme

(43). Their goal is to pursue the use of synthetic mtDNA to treat diseases associated with mtDNA mutations.

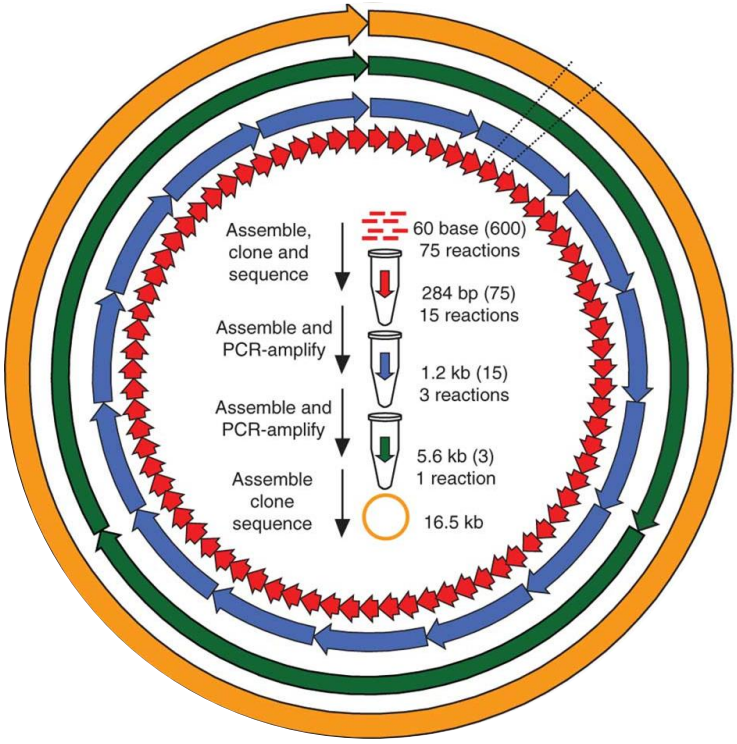


Figure 6. Assembly of the Synthetic Mouse MtGenome by Four Sub-Assembly Steps.

The 60-base oligonucleotides were assembled into 284bp segments (red arrows). These were then assembled into 1.2kb segments (blue arrows), which were assembled into 5.6kb segments (green arrows). The 5.6kb segments were combined to form the complete genome along with the 221bp repeat regions and the pUC19 vector. (Figure modified from Gibson *et al.* (43))

Problem

Efforts to evaluate human samples for the presence of mtDNA deletions have been complicated by non-specific background bands. These background bands are present in amplifications of samples presumed to be deletion-free, and therefore, optimization of this method is needed. Because the sequence of the chemically synthesized mtGenome is known and is easily harvested from bacterial cells in the absence of autosomal mammalian DNA, it provides a previously unavailable control to evaluate common molecular techniques. It is of a known size, absent of genetically engineered deletions, and can therefore be used to optimize the long range (LR) PCR protocol. Any background bands will be presumed to be PCR artifacts and not evidence of low level deletion products. The genome may also be used to evaluate the quality of data produced by different sequencing platforms. The synthetic genome was synthesized according to the fully sequenced *Mus musculus* genome (GenBank record NC_005089).

Hypothesis: A synthetic genome can be used to optimize and evaluate common molecular techniques

Specific Aims:

Aim 1- Optimize LR PCR protocol using the mouse synthetic genome and human samples of a known state.

A: Optimize parameters for the LR PCR protocol in order to amplify greater than 10kb mtDNA without non-specific background products

i. Evaluate using Sybr® Gold (Life Technologies) staining.

ii. Evaluate using ethidium bromide staining.

B: Verify optimized procedure using previously identified samples with intact mtGenomes and those with deletions greater than 1kb.

C: Conduct Sanger and next generation sequencing (MiSeq System and Ion PGM™ System) of the synthetic mtDNA.

i. Evaluate overall quality of the sequence data generated by the different methods.

ii. Because the exact base composition of the synthetic genome is known, any errors will be identified through comparison to the *Mus musculus* reference genome. Identify commonalities in error introduction. Identify errors that are associated with specific motifs and homopolymeric stretches. Compare and contrast between methodologies.

Aim 2- Identify mtDNA deletions in human samples provided by Texas Alzheimer's Research and Care Consortium.

A: Screen samples using the optimized LR PCR protocol to identify samples with deletions and determine the approximate size of the deletion.

CHAPTER 2

MATERIALS AND METHODS

Synthetic Mouse MtGenome

A material transfer agreement has been established between the University of North Texas Health Science Center and the J. Craig Venter Institute to obtain an *E. coli* strain containing a plasmid-borne synthetic mouse mtGenome. The sample was streaked onto Luria broth agar plates containing 12.5µg/mL chloramphenicol. Liquid cultures were prepared from isolated colonies and incubated with aeration at 37°C overnight. Five 0.5mL aliquots of liquid culture in 80% sterilized glycerol were stored at -80°C.

Frozen stock was used to inoculate 200mL of Luria broth containing 12.5µg/mL chloramphenicol and incubated at 37°C overnight. Following slight agitation, 160mL of culture was removed and replaced with 160mL of fresh Luria broth. The culture was incubated an additional 30 min at 37°C with shaking. Cells were induced to produce a high copy number of recombinant plasmid by addition of 200µL of 10X CopyControl™ Induction Solution (Epicentre® Biotechnologies) and an additional two hours of 37°C incubation with shaking.

To harvest the plasmid DNA, 50mL of culture was transferred to each of two 50mL conical tubes and centrifuged at 1500xg for 5 min. The supernatant was removed, and this step was repeated. Plasmid was harvested using the PerfectPrep™ EndoFree Maxi Kit (5 PRIME, Gaithersburg, MD) according to the manufacturer's instructions. DNA concentration was measured using both the NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA) and Qubit® dsDNA BR Assay (Life Technologies). This concentration was

measured to be 180ng/µL using the NanoDrop 2000 UV-Vis Spectrophotometer, and 8ng/µL using the Qubit® dsDNA BR Assay. This discrepancy will be discussed in Chapter 3.

Human DNA Samples

Three different human DNA sample types were used in this study. Control A is DNA isolated from peripheral whole blood of a healthy 65 year old donor. In addition, cell line DNA and TARCC samples were used.

Cell Lines: Human DNA from two cell lines was utilized. The first consists of DNA extracted from a cybrid cell line harboring a large deletion associated with Kearns-Sayre Syndrome (KSS). This cell line contains a 7.5kb deletion between positions 7,982 and 15,504 in approximately 60% of its mtGenomes (44). The second control is expected to be free of deletions and consists of DNA extracted from an HL-60 (HL60) cell line (ATCC® CCL-240™)(ATCC, Manassas, VA).

TARCC: Seven TARCC (Texas Alzheimer's Research and Care Consortium) samples harboring deletions were assayed. Sample DNA was previously extracted from peripheral blood buffy coats. The presence and proportion of deletion-containing to non-deletion-containing mtDNA were also previously determined using a real-time quantitative PCR assay as described by Phillips *et al.* (45) (Table 2). This assay measures two targets on the mtGenome: one target is in a region where deletions commonly occur and the second target is in a conserved region.

Table 2. TARCC Samples.

The proportion of deletion-containing to intact mtDNA from TARCC samples was measured using a previously described qPCR assay (45).

Sample	Deletion Proportion (%)
T09	28
T35	16
T94	10
T162	20
T168	10
T180	11
T190	12

Long Range PCR Amplifications

Previously published primer sets were utilized for amplification of human mtDNA (Table 3) and checked for specificity against the revised Cambridge reference sequence (rCRS) published under Genbank accession number NC_012920. Primers to amplify the synthetic mouse mtGenome (Table 3) were designed using the Primer Design tool from the National Center for Biotechnology Information (NCBI) and based on the *Mus musculus* sequence published under Genbank accession number NC_005089. The open-source Primer Design software Oligoanalyzer 3.1 (Integrated DNA Technologies, Coralville, IA) was used to assess potential primer pairs for melting temperature compatibility, and structural deficiencies including self-dimerization, heterodimerization, and hairpins. Sequence specificity was assessed using the BLAST tool available through NCBI(46)(45).

Table 3. Primer Sets Utilized in LR PCR Amplification.

Primer Set	Sequence (5' to 3')	Specificity	Intact Amplicon	Source
A4 (Forward)	CCCCATGCTTACAAGCAAGT	Human	16kb	(46)
22R (Reverse)	AGCTTTGGGTGCTAATGGTG			(46, 47)
161MitoF (Forward)	TCGCACCTACGTTCAATATTACAGGCG	Human	16.4kb	(15)
16510MitoR (Reverse)	TAGGAACCAGATGTCGGATACAGTTC			(15)
Mus2066F (Forward)	ATGAACGGCTAAACGAGGGTCCAAC	Mouse	12.7kb	*in silico* design
Mus14745R (Reverse)	GGAGGAAGAGGAGGTGAACGATTGC			

Initially, two commercially available amplification kits were tested: the SequelPrep™ Long PCR Kit with dNTPs (Invitrogen) and the PrimeSTAR® GXL DNA Polymerase reagent set (TaKaRa Bio USA). The SequelPrep™ assay reactions consisted of 1X SequelPrep™ Reaction Buffer, 0.4µL dimethyl sulfoxide, 0.5X SequelPrep™ Enhancer A, 1.8 Units of SequelPrep™ Long Polymerase, 0.5µM each of forward and reverse primers, template DNA, and DNase-free water for a final reaction volume of 20µL. The PrimeSTAR® GXL assay reactions consisted of 1X PrimeSTAR® GXL Buffer, dNTP Mixture (200µM of each dNTP), forward and reverse primers (0.2µM each), 1.25 Units of PrimeSTAR® GXL DNA Polymerase, template DNA, and DNase-free water for a final reaction volume of 50µL. The thermal cycling parameters are detailed in Table 4.

Table 4. Thermal Cycling Parameters for SequelPrep™ and PrimeSTAR® GXL Assays.

SequelPrep™ Assay			PrimeSTAR® GXL Assay
A4/22R	161MitoF/16510MitoR	Mus2066F/Mus14745R	All Primer Sets
94°C 2:00 10 Cycles of: 94°C 0:10 50.4°C 0:30 68°C 16:00 20 Cycles of: 94°C 0:10 50.4°C 0:30 sec 68°C 22:40 Final Extension: 72°C 5:00	94°C 2:00 10 Cycles of: 94°C 0:10 57°C 0:30 68°C 16:00 20 Cycles of: 94°C 0:10 57°C 0:30 68°C 22:40 Final Extension: 72°C 5:00	94°C 2:00 10 Cycles of: 94°C 0:10 60°C 0:30 68°C 13:00 20 Cycles of: 94°C 0:10 60°C 0:30 68°C 19:40 Final Extension: 72°C 5:00	30 Cycles of: 98°C 0:10 68°C 10:00

Amplification products were analyzed alongside a 1kb Plus DNA Ladder (Life Technologies) using a 0.8% agarose gel (100 volts, 30min to 90min). Products were visualized using 1X SYBR® Gold and/or ethidium bromide staining. Assay performance was assessed by visual analysis of the stained amplification products. Optimization experiments included variations in DNA template concentrations, cycle number, annealing/extension times, and annealing/extension temperatures as detailed in Chapter 3.

Sequencing of Synthetic Mouse MtGenome

Sequencing of the synthetic mouse mtGenome was conducted utilizing Sanger sequencing, the MiSeq System, and the Ion PGM™ System.

Sanger Sequencing

The PrimeSTAR® GXL assay and Mus2066F/Mus14745R primer set (Table 3) were used to amplify an approximately 12.7kb fragment of the mtGenome. Amplicons were purified using ExoSAP-IT® (Affymetrix, Santa Clara, CA). Twenty microliters of ExoSAP-IT® reagent

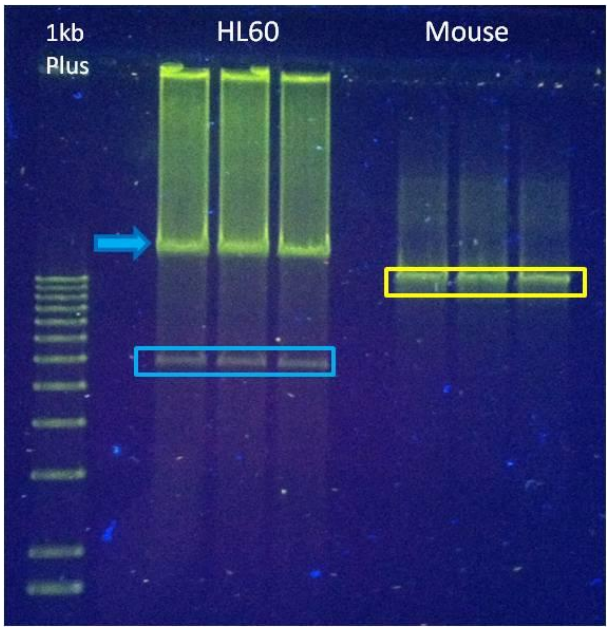
was added to the 50µL reaction. The samples were incubated at 37°C for 15min followed by 80°C for 15min. Cycle sequencing PCR was conducted in the forward and reverse directions. Reactions were prepared using the ABI BigDye® Terminator™ v1.1 Cycle Sequencing Kit (Life Technologies) and consisted of 1.5µL 3.3µM primer, 1µL BigDye® Terminator v1.1, 5µL Sequencing Buffer, 1µL of PCR product, and 6.5µL DNase-free water. Samples were amplified with the following thermal cycling parameters: 96°C for 3min followed by 25 cycles of 96°C for 15sec, 50°C for 10sec, and 60°C for 3min. Reactions were purified using the BigDye® XTerminator™ Purification Kit (Life Technologies) by addition of 27.5µL DNase-free water, 22.5µL SAM™ Solution, and 5µL BigDye® XTerminator™ and shaking at 2500rpm for 30min. Samples were analyzed using capillary electrophoresis on the 3130*xl* Genetic Analyzer (Life Technologies). Data were analyzed using Sequence Scanner v1.0 (Life Technologies) and compared to the mouse mtGenome sequence published under GenBank record NC_005089.

MiSeq System

Three samples were prepared for sequencing on the MiSeq System: 1) the unamplified synthetic mouse mtGenome prepared as described above under "Synthetic mtGenome"; 2) a gel purified amplicon from a 1,000-fold dilution of the synthetic mouse mtGenome amplified with the PrimeSTAR® GXL assay and Mus2066F/Mus14745R primer set (Table 3); and, 3) a gel purified amplicon of the background band co-amplified with the HL60 DNA using the PrimeSTAR® GXL assay and the 161MitoF/16510MitoR primer set (Table 3).

In order to isolate the PCR amplicons, the entire PCR product was loaded in three equal aliquots into wells of a 0.8% agarose gel. Voltage was applied for five minutes at 100 volts then reduced to 75 volts for 2hrs. The gel was stained with Sybr® Gold and the product was

visualized on an ultraviolet light box (Figure 7A). The gel containing the amplicon was then excised with a sterile scalpel, and the amplicon was extracted from the gel using the Midi FlexTube kit (IBI scientific, Peosta, IA). Following ethanol precipitation, the amplicon was suspended in DNase-free water. Successful purification of the amplicons was confirmed using gel electrophoresis (Figure 7B).

A) 

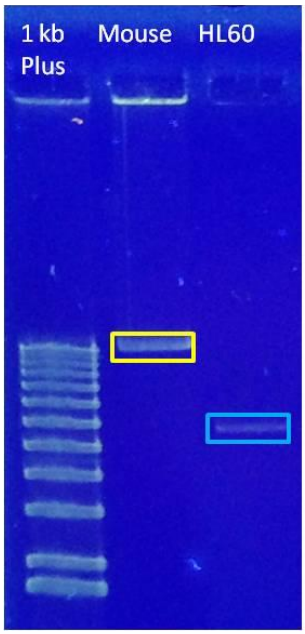
B) 

Figure 7. Gel Extraction and Purification.

Lane 1 is a 1kb Plus DNA Ladder. (A) The HL60 and synthetic mouse mtGenome samples were amplified using the PrimeSTAR® GXL assay. For each, the entire 50µL reaction was loaded in equal aliquots into three wells of a 0.8% agarose gel. The intact amplicon is visible for the HL60 sample (blue arrow). The HL60 background band (blue box) and the syntetic mouse amplicon (yellow box) were excised for sequencing. (B) Successful purification of the amplicons was verified using gel electrophoresis: HL60 and mouse amplicons shown with blue and yellow boxes, respectively. Note: The gel pictured in (A) was run at a lower voltage and for more time than the gel pictured in (B).

DNA quantification was performed using the Qubit® dsDNA BR Assay Kit (Life Technologies). The unamplified synthetic mouse mtGenome and the gel-purified synthetic mouse amplicon were measured at 8ng/µL and 2.2ng/µL, respectively. The HL60 background band was not sufficiently concentrated to be quantified with the Qubit® assay. However, since the purified amplicon was visualized in the gel (Figure 7B), the sample was prepared for sequencing. The mouse samples were normalized to 1ng/µL. For each sample, 5µL was used as input for preparation using the Nextera® XT Sample Prep Kit (Illumina). Samples were sequenced using the MiSeq System. Sequence reads were aligned to the Genbank sequence NC_005089 and NC_012920 for the mouse and human sequences, respectively, using NextGENe® software (SoftGenetics, State College, PA).

Ion PGM™ System

Two samples were prepared for sequencing on the Ion PGM™ System: 1) the unamplified synthetic mouse mtGenome prepared as described above under "Synthetic mtGenome"; 2) A 10,000-fold dilution of the synthetic mouse mtGenome amplified with the PrimeSTAR® GXL assay and Mus2066/Mus14745R primer set (Table 3). The LR PCR amplicon was purified using ExoSAP-IT® as described above under "Sanger Sequencing". DNA quantifications were performed using the Qubit® dsDNA BR Assay Kit. The unamplified synthetic mouse mtGenome and purified synthetic mouse amplicon were measured at 8ng/µL and 10ng/µL, respectively.

For both samples, 100ng was used as input for preparation using the NEBNEXT® Fast DNA Fragmentation & Library Prep Set (New England Biolabs) and NEXTflex™ DNA

Barcodes (Bioo, Scientific Corp, Austin TX). The DNA was amplified through emulsion PCR on the Ion One Touch™ 2 System (Life Technologies) using the Ion PGM™ Template OT2 400 Kit (Life Technologies) for preparation of 400bp read libraries. Sequencing was conducted using the Ion PGM™ Sequencing 400 Kit (Life Technologies) and on the Ion 318™ Chip Kit v2 (Life Technologies). NextGENe® software (SoftGenetics) was utilized for data analysis.

CHAPTER 3

RESULTS

LR Results

Initial LR PCR attempts using the SequelPrep™ assay resulted in multiple non-reproducible background bands (Figure 2). In this experiment, mtDNA from three healthy donors was amplified in triplicate using the A4/22R primer set (Table 3). The expected intact amplicon was observed in the replicates for each sample (Figure 2, aligned with blue arrow). However, also observed in the replicates were multiple faint background bands (Figure 2, yellow asterisks). These bands were not consistent in size or number within replicates of the same sample. Reproducibility of smaller bands would be expected if these products were the result of mtDNA deletions present within the sample. In order to establish this phenomena as recurring, amplification was conducted using the same primers (A4 and 22R) and assay conditions for two DNA samples not expected to contain mtDNA deletions, Control A and HL60 (Figure 8).

Control A consists of DNA extracted from whole blood of a healthy 65 year old male donor. The second control is extracted DNA from an HL60 cell line obtained from ATCC. For both samples, the intact 16kb amplicon was detected (Figure 8, aligned with blue arrow), along with multiple background bands (Figure 8, yellow asterisks). Most of the background bands were similar in size between the two amplifications. Given the previously observed non-reproducibility of the background bands among replicates of a single sample, the similarity in background bands observed here was unexpected. These samples were not expected to contain mtDNA deletions, so it was unlikely that the six background products observed for each sample

represented true mtDNA deletions. A more likely explanation was that these background bands were the result of non-specific primer binding. This would also explain why the majority of background bands were observed in the same location between the two samples. If the thermal cycling conditions allowed for amplification of areas where the primers had partial homology elsewhere in the mtGenome, it would be expected that these non-specific amplification products would be similar between samples.

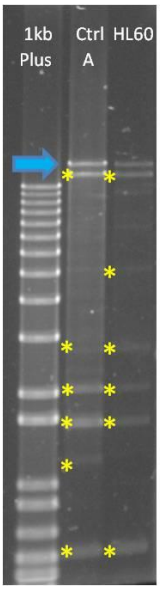


Figure 8. SequelPrep™ Amplifications.

LR PCR of Control A (CtrlA) and HL60 samples expected to be free of deletions. Lane 1 is a 1kb Plus DNA Ladder. The main, intact product is present in each and is the largest fragment (blue arrow). Multiple background bands are visible (yellow asterisks), the majority of which are of similar size between samples.

In order to evaluate whether increased annealing temperature could eliminate background bands resulting from non-specific primer binding, an annealing temperature gradient experiment was conducted. Six reactions each were prepared using the Control A and HL60 samples. Annealing temperatures ranged from 53°C to 58°C and increased in 1°C increments between

reactions. While the overall number of background bands decreased as annealing temperature increased, the position of these bands was not consistent between amplifications (Figure 9). Also, the intensity of the intended product decreased with increasing annealing temperature, indicating inefficient amplification at higher annealing temperatures. It is possible that additional background products are still produced at these higher annealing temperatures, but the detection method is not sensitive enough for them to be visualized.

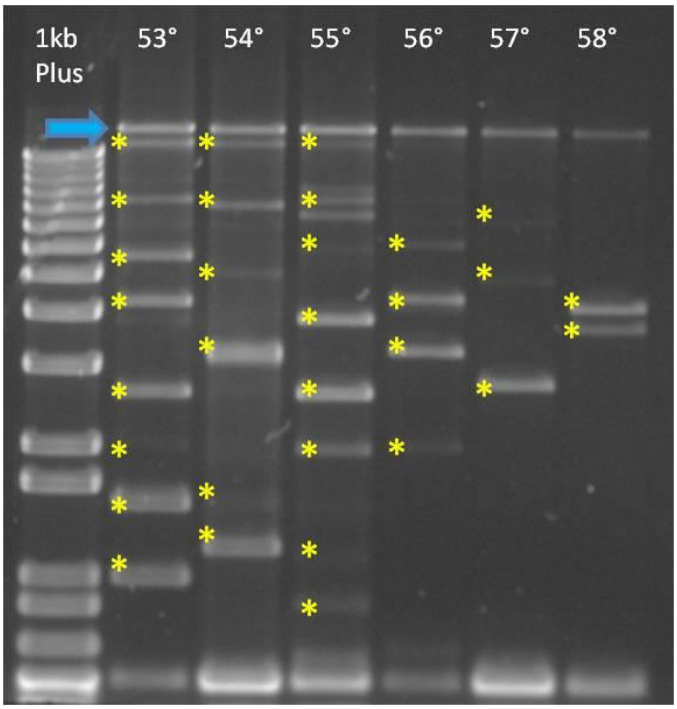


Figure 9. SequelPrep™ Assay Annealing Temperature Gradient Experiment.

Amplifications of HL60 DNA. Lane 1 is a 1kb Plus DNA Ladder. Annealing temperature is indicated at the top of the lane. The intact product is present in each and is the largest fragment (aligned with blue arrow). Multiple background bands are visible (yellow asterisks) which are inconsistent between replicates and decrease in number with increased annealing temperature.

Because of the large number and random nature of the background bands, no further optimization using the SequelPrep™ assay was pursued. Instead, amplification was conducted using the PrimeSTAR® GXL DNA Polymerase assay. This chemistry utilizes different thermal cycling parameters than the SequelPrep™ assay (Table 3), and recommends using longer oligonucleotides with a higher annealing temperature than is characteristic of A4 and 22R. Therefore, amplification was performed using a previously published primer set from Khrapko *et al* (Table 3).

As a preliminary step, the PrimeSTAR® GXL assay was used for amplification of Control A and HL60 along with the KSS sample which contains a known deletion. The KSS sample consists of DNA extracted from a cybrid cell line in which approximately 60% of mtGenomes harbor a deletion from positions 7982 to 15,504 that was previously reported to be present in patients with Kearns-Sayre syndrome (44). The mtDNA concentrations for the Control A, HL60, and KSS samples were measured to be 14.8pg/µL, 1.9pg/µL and 33.0pg/µL, respectively. The samples were amplified in duplicate. For the HL60 and KSS samples, the first reaction was prepared using 1µL of sample template, and the second reaction using 2µL of template. For the Control A sample, the first reaction contained 6pg while the second reaction contained 14.8pg. For each -sample, the amplification with less DNA input was successful. The deletion-containing amplicon was successfully detected in the KSS sample (Figure 10, aligned with green arrow) and the intact amplicon was detected in the Control A and HL60 samples (Figure 10, aligned with blue arrow). For the KSS sample, the intact amplicon was only very faintly detectable. Few background bands were visible (Figure 10, yellow asterisks).It is unclear why amplifications containing relatively higher amounts of input DNA failed. These reactions produced a large amount of smearing with some indistinct banding and no detection of the intact

or deletion-containing amplicons (Figure 10, yellow stars). Given the range of input DNA in successful amplifications (6pg to 33pg) compared with the failed amplifications (3.8 to 66pg), it is unlikely that this was a result of too much DNA template.

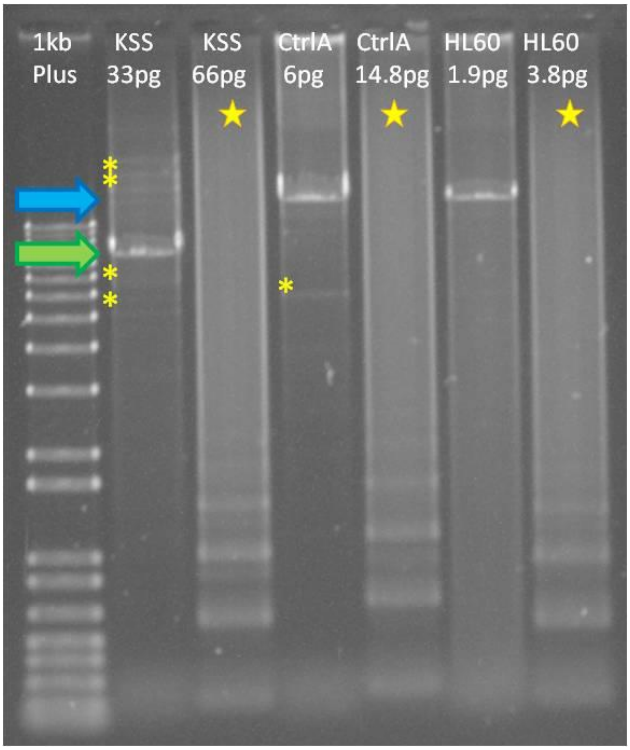


Figure 10. PrimeSTAR® GXL Assay.

Amplification of KSS, Control A (CtrlA), and HL60 DNA. Lane 1 is a 1kb Plus DNA Ladder. The truncated, deletion-containing amplicon is present for the KSS 33pg amplification (green arrow), while the intact amplicon is present in Control A 5.98pg and HL60 1.9pg amplifications (aligned with blue arrow). Few background bands are visible (yellow asterisks). Amplifications at higher DNA concentrations failed to amplify intact or deletion-containing products (yellow stars).

Amplification was also performed with the SequelPrep™ assay using the 161MitoF and 16510MitoR primer set with the same samples and template concentrations. Amplification was successful for all reactions with the intact amplicon visible in the Control A and HL60 amplifications (Figure 11, aligned with blue arrow) and the deletion-containing amplicon visible for the KSS amplification (Figure 11, aligned with green arrow). Use of this alternative primer set reduced the number of background products observed and bands were consistent between replicates of the same sample (Figure 11, yellow asterisks). This suggested that background bands may have been the result of non-specific primer binding. Given that fewer bands were observed, it was hypothesized that utilizing this primer set with a higher annealing temperature would be more successful in eliminating background bands.

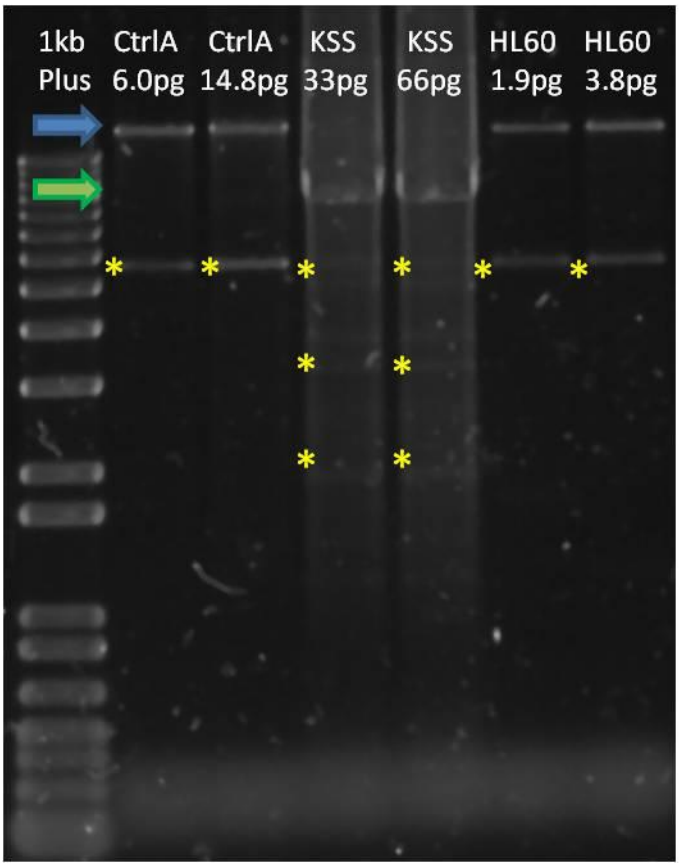


Figure 11. SequelPrep™ Assay Amplification Using 161MitoF and 16510MitoR Primer Set.

Amplification of KSS, Control A, and HL60 DNA. Lane 1 is a 1kb Plus DNA Ladder. The intact amplicon is present in the Control A (CtrlA) and HL60 amplifications (aligned with blue arrow), and the deletion-containing amplicon is present in the KSS amplifications (aligned with green arrow). Additional products are indicated by yellow asterisks.

To address that hypothesis, a second annealing temperature gradient experiment was conducted. Reactions were prepared using HL60 DNA at 1.9pg/reaction and 3.8pg/reaction. The annealing temperature was increased in 3°C increments ranging from 57°C to 69°C. The intact amplicon is visible for both DNA concentrations and at all annealing temperatures tested. As with the previous gradient experiment, the intact amplicon becomes fainter with increasing temperature indicating an overall decrease in amplification efficiency.

In general, reactions with 3.8pg DNA resulted in a larger number of background bands (Figure 12, yellow asterisks). With the exception of the 63°C reactions, the intensity of the background bands observed also decreased with increasing temperature. Due to a programming error, the reactions with an annealing temperature of 63°C underwent an additional five cycles of amplification (a total of 35 cycles). This resulted in background bands that were markedly more pronounced yet the intact amplicon is not visibly more intense than what is observed in the other reactions. This indicates that increasing the cycle number results in preferential amplification of the background bands over the intact amplicon. This also supports the previous hypothesis that background bands are still present at higher annealing temperatures but the detection method is not sensitive enough for visualization.

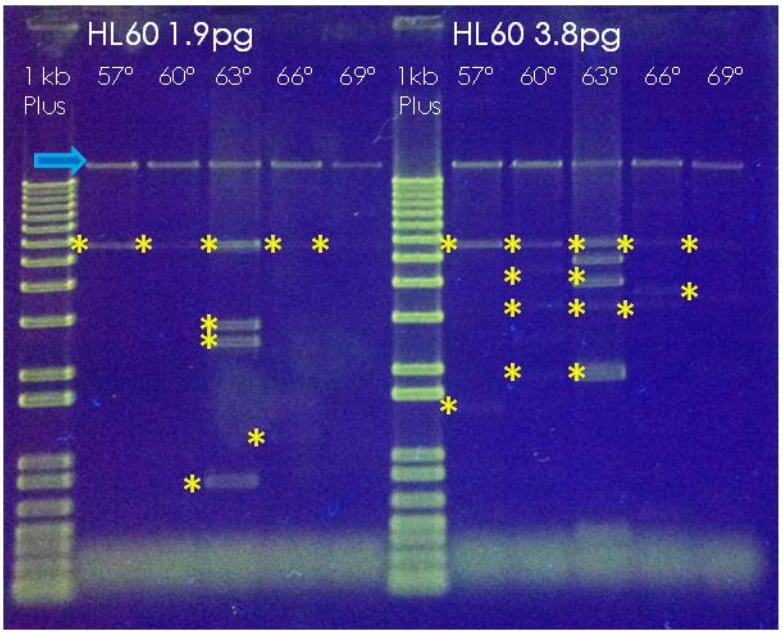


Figure 12. SequelPrep™ Assay Annealing Temperature Gradient Using Primer Set 161MitoF and 16510MitoR.

Amplifications of HL60 at 1.9pg per reaction and 3.8pg per reaction. Lane 1 is a 1kb Plus DNA Ladder. Due to a programming error, the reactions with an annealing temperature of 63°C underwent an additional five cycles of amplification. The intact amplicon was detected at all annealing temperatures for both DNA concentrations (aligned with blue arrow). Background bands are indicated by yellow asterisks. Image is shown in color due to better resolution for visualization of background bands.

Thus far, optimization efforts were successful in reducing the number and intensity of background bands, but were unable to eliminate them entirely. While Control A and HL60 were assumed to be free of mtDNA deletions, this assumption was not experimentally verified due to spurious background bands. Therefore, the synthetic Mus musculus mtGenome, the first synthetic genome of an organelle and known to be free of mtDNA deletions (43), was utilized in further optimization experiments. The genome was synthesized as described in Figure 6 and is based on the 16.3kb *Mus musculus* genome published under GenBank record NC_005089.

The synthetic genome was harvested from recombinant *E.coli* cells using the PerfectPrep™ Endofree Maxi kit in two aliquots. The concentration of the DNA extract was measured at 180ng/µL using the NanoDrop 2000UV-Vis Spectrophotometer and at 8ng/µL using the Qubit® dsDNA BR Assay. The exact number of molecules and/or concentration of these extracts is not known due to the incongruent measurements produced by these two methods as well as the mere non-specificity of the measurement assays. Therefore, LR PCR amplification was conducted using either 1µL or 2µL of template for both extraction aliquots. Amplification was conducted using a newly designed primer set (Table 3) and and the SequelPrep™ and PrimeSTAR® GXL assays (Figure 13). In all amplifications, the expected 12.7kb amplicon was detected (Figure 13, aligned with blue arrow); however, the band was not discreet and all amplifications exhibited a large amount of smearing. These results were more pronounced in the SequelPrep™ reactions.

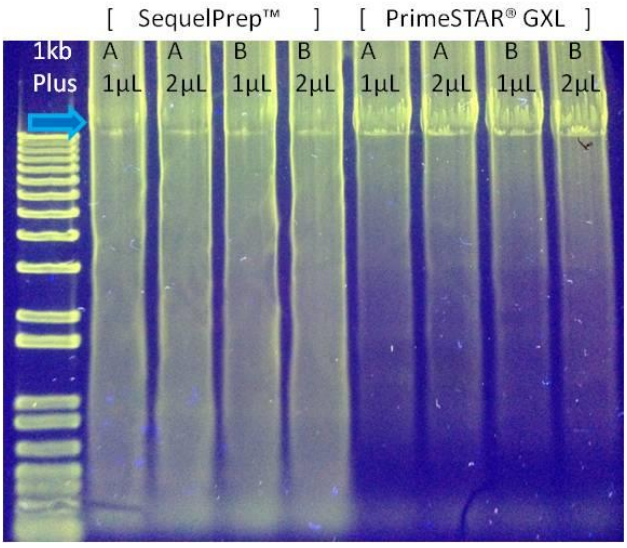


Figure 13. SequelPrep™ and PrimeSTAR® GXL Assays Amplification of the Synthetic Mouse MtGenome.

Lane 1 is a 1kb Plus DNA Ladder. Amplification was conducted using both LR PCR assays and isolate aliquots of the synthetic mouse mtGenome (A and B) with either 1µL or 2µL template. The intended product was observed (aligned with blue arrow) but the band was not discreet. Excessive smearing indicates that the reactions may have been overloaded with too much template. Smearing makes it difficult to discern the individual background bands.

Because of the relatively poor performance of the SequelPrep™ assay relative to the PrimeSTAR® GXL assay, further optimization efforts focus solely on the PrimeSTAR® GXL assay. Given the large disparity of DNA quantification results from the NanoDrop 2000 and Qubit® dsDNA BR Assay, the quantity of the DNA template added to the amplification was not known. By either measurement, the DNA quantity was much greater than used in previous amplifications. Therefore, the following experiment addressed two hypotheses. The first hypothesis was that smearing was the result of non-amplified DNA forced through the gel. The

second was that the smearing was a result of excess template and amplification would be more efficient if less DNA was added.

In order to address the hypothesis that smearing was the result of unamplified template, a single amplification reaction was prepared using 1µL of undiluted synthetic mouse mtDNA without the addition of the primers. No band or smearing was observed in the reaction without primers indicating that both are the result of amplification. To address the hypothesis of excess DNA, reactions were prepared that contained 1µL of undiluted template along with reactions containing 1µL of 0.5, 0.25, and 1.25×10^{-2} dilutions. No notable differences were observed among amplifications of non-diluted and diluted DNA samples. All amplifications exhibited smearing (Figure 14).

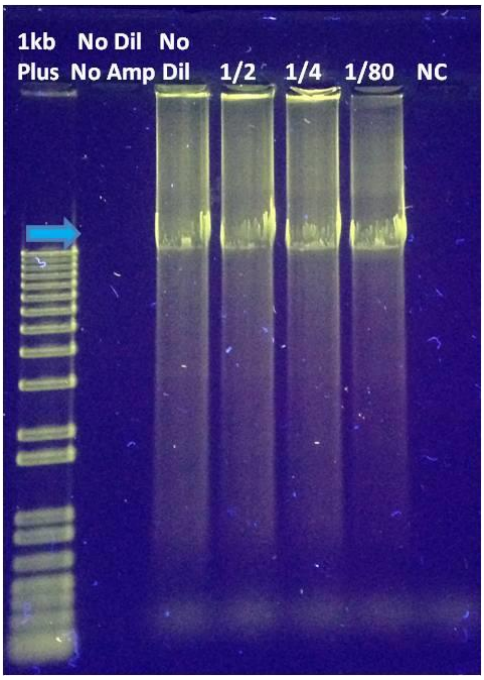


Figure 14. PrimeSTAR® GXL Amplification of the Synthetic Mouse MtGenome Dilutions.

Lane 1 is a 1kb Plus DNA Ladder. No smearing was observed of the undiluted, unamplified sample (No Dil No Amp). Amplification of undiluted (No Dil) and diluted DNA (0.5, 0.25, and 1.25×10^{-2} labeled as 1/2, 1/4 and 1/80) produced the same smeared product (aligned with blue arrow) with no observable difference among dilutions. NC is a PCR negative control.

To further address the smearing observed in Figure 14, two experiments were performed. The first experiment tested whether smearing was the result of inefficient primer binding due to the two-step thermal cycling parameters. The 0.25 and 1.25×10^{-2} dilutions were amplified using a three-step thermal cycling program that included a 15 second annealing step at 65°C. The second experiment tested whether the reactions were still overloaded with template at a 1.25×10^{-2} dilution. Therefore, the dilution series was continued to include dilutions of 1.25×10^{-3}, 1.25×10^{-4}, and 1.25×10^{-5}. The intact amplicon is present for all amplifications regardless of

template concentration or thermal cycling parameters (Figure 15, aligned with blue arrow). A single background band was observed for the 0.25 and 1.25×10^{-2} dilution reactions for both sets of thermal cycling conditions (Figure 15, yellow asterisks), but was not observed in the more dilute samples. They resulted in improved resolution with the most discreet band observed for the 1.25×10^{-5} dilution. Regardless of which assay is more accurate, the concentration measured using the NanoDrop 2000 UV-Vis Spectrophotometer and the Qubit® dsDNA BR Assay corresponds to 2.25pg and 0.1pg, respectively. As resolution improved, a second, faint band was observed directly beneath the intact amplicon (Figure 15, aligned with red arrow).

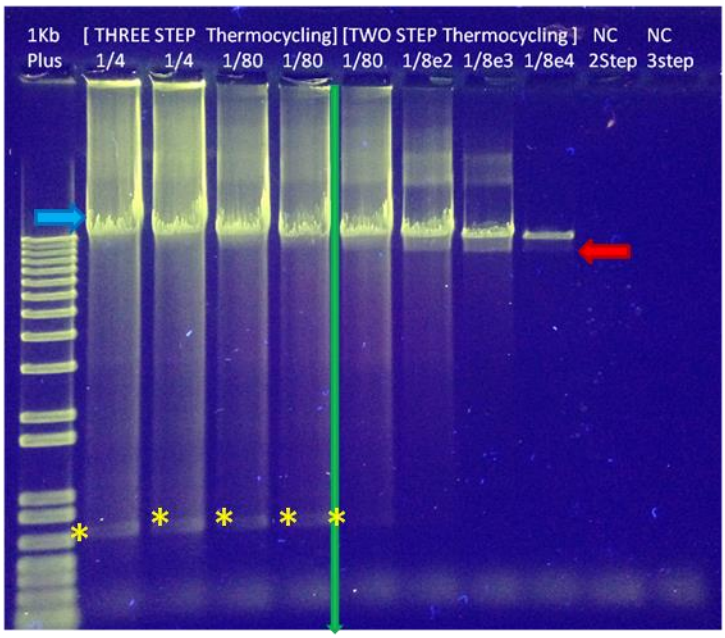


Figure 15. PrimeSTAR® GXL Amplification of the Synthetic Mouse MtGenome.

Lane 1 is a 1kb Plus DNA Ladder. The first four samples (left of the vertical green line) were amplified using a three-step thermal cycling program. The dilution series was continued for two-step amplification of template (right of the vertical green line). The intact amplicon is present for all (aligned with blue arrow). The yellow asterisks indicate the presence of a single background band. Further dilutions resulted in improved resolution with the most discreet band observed for the 1.25×10^{-5} (or 1/8e4) dilution. As resolution improved, a second, faint band was observed directly beneath the intact amplicon (aligned with red arrow). NC 2step and NC 3step are PCR negative controls.

To briefly summarize the results for the PrimeSTAR® GXL assay thus far, amplification yielded the intact amplicon with few background bands for both synthetic mouse mtGenome (Figure 15) and human (Figure 10) samples. In order to evaluate whether further dilutions would result in elimination of the single background band observed in amplifications of the synthetic mouse mtGenome (Figure 15, aligned with red arrow), the dilution series was continued to 1.25×10^{-6} and 1.25×10^{-7}. Amplification of the synthetic mouse mtGenome at a dilution of

1.25×10^{-6} yielded only a very faintly visible amplicon (Figure 16, aligned with blue arrow), and no product was observed at a dilution of 1.25×10^{-7}. In order to determine if the PrimeSTAR® GXL assay was able to detect mtDNA deletions in human samples, two TARCC samples were amplified. These samples, identified as T180 and T190, were previously tested with a qPCR assay, and determined to have deletion-containing mtDNA at a ratio of 11% and 12%. The HL60 and KSS samples were amplified as controls for the detection of intact and deletion-containing amplicons, respectively. For the HL60 and TARCC samples, the intact amplicon was observed (Figure 16, aligned with green arrow). The KSS sample yielded the expected deletion-containing amplicon (Figure 16, aligned with yellow arrow). For all human samples, a single secondary product was observed (Figure 16, aligned with red arrow).

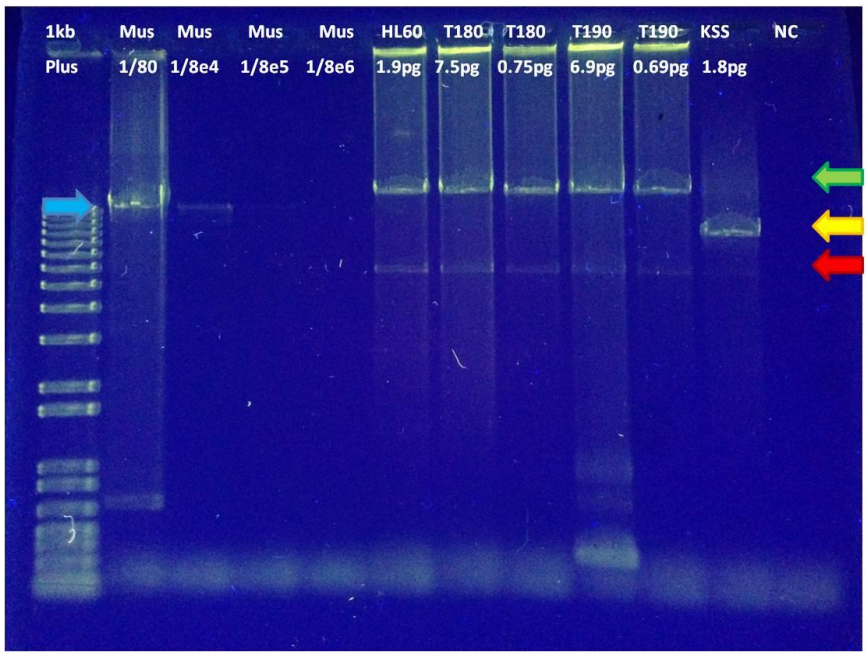


Figure 16. PrimeSTAR® GXL Amplification of the Synthetic Mouse MtGenome and Human Samples.

Lane 1 is a 1kb Plus DNA Ladder. The mouse amplicon (Mus) was observed at 1.25×10^{-2} (or 1/80) and 1.25×10^{-5} (or 1/8e4) dilutions, and only very faintly observed at 1.25×10^{-6} (or 1/8e5) aligned with blue arrow). The intact amplicon (aligned with green arrow) was observed for the HL60 and TARCC samples (T180 and T190), and the deletion-containing amplicon (aligned with yellow arrow) was observed for the KSS sample. For all human samples, a single secondary band was observed (aligned with red arrow). NC is a PCR negative control.

To evaluate the efficacy of the two dyes, Sybr® Gold and ethidium bromide, the synthetic mouse mtGenome was amplified at dilutions of 1×10^{-3} and 1×10^{-4} and visualized with both dyes. Sybr® Gold is the more sensitive dye. While the main amplicon is less intense, a faint signal below this amplicon is more visible with the Sybr® Gold (Figure 17).

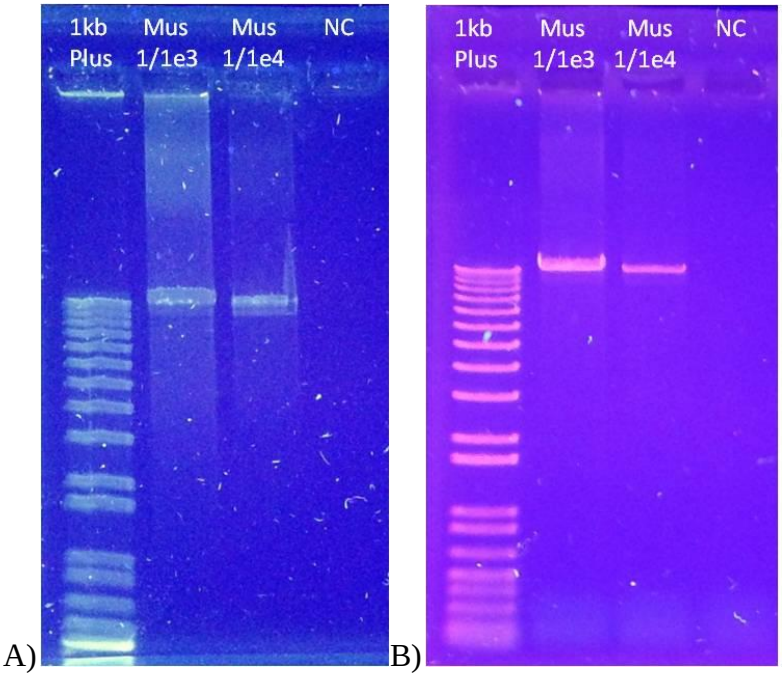


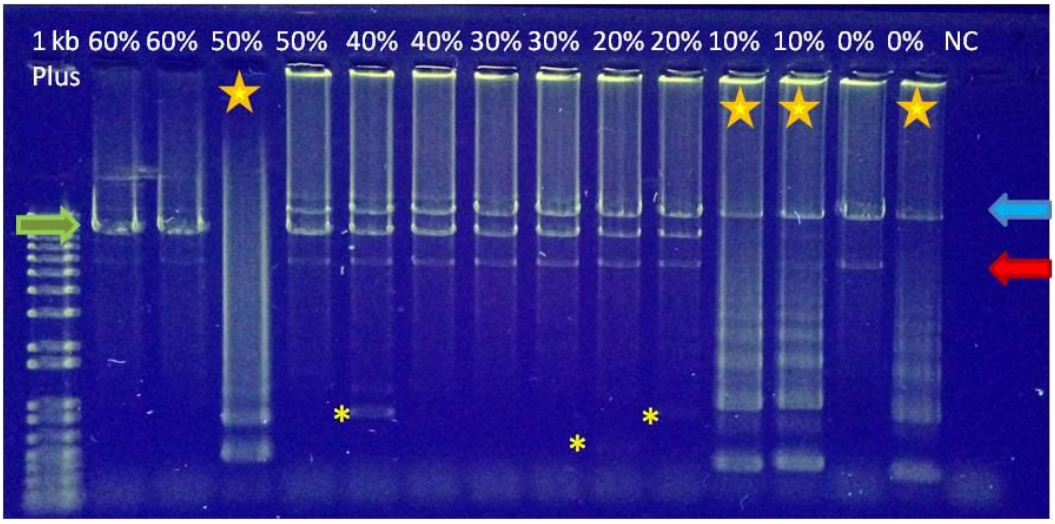
Figure 17. Comparison of Nucleic Acid Dyes.

Products were visualized with Sybr® Gold (A) and ethidium bromide (B). Lane 1 is a 1kb Plus DNA Ladder. For the 1×10^{-3} (Mus 1/1e3) and 1.25×10^{-4} (Mus $1/1e^{4}$) dilutions of the synthetic mouse mtGenome, the intact amplicon is visible using both dyes. Sybr® Gold resulted in clearer observation of a secondary signal directly beneath the main amplicon. NC is a PCR negative control.

Because deletion products were not observed as expected for the TARCC samples T180 and T190, an experiment was conducted to determine the minimum ratio at which the assay could be used to detect deletions. As previously stated, the KSS sample contains approximately 60% of mtGenomes harboring an 8kb deletion while the HL60 sample is not known to contain deletions. Therefore, a series of mixtures was prepared of KSS and HL60 DNA in order to test for the presence of the deletion-containing amplicon at a decreasing ratio to the intact amplicon. HL60 DNA was added in order to decrease the deletion-containing proportion in 10% increments from 60% to 0%.

Products were visualized using both SYBR® Gold (Figure 18A) and ethidium bromide (Figure 17B). The deletion-containing amplicon was readily observed at ratios ranging from 60% down to 20%, but only in one of the 10% duplicate reactions (Figure 18A). It is likely that the deletion-containing product was obscured by noticeable smearing and blurred banding that occurred in both 10% reactions along with one of the 50% and 0% reactions each (Figure 18, yellow stars). Duplicate reactions were prepared from a single master mix and then dispensed into separate reaction tubes, so this phenomena was unlikely to be the result of pipetting error. The single background band observed in previous amplifications of human DNA (Figure 16, aligned with red arrow) was observed here in all reactions except those with the smearing (Figure 18, aligned with red arrow). An additional background band was visible in a single 40% reaction (Figure 18, yellow asterisks). However, background banding was less visible using ethidium bromide staining, and both 20% reactions exhibited additional background only visible using SYBR® Gold (Figure 18A, yellow asterisks).

A)



B)

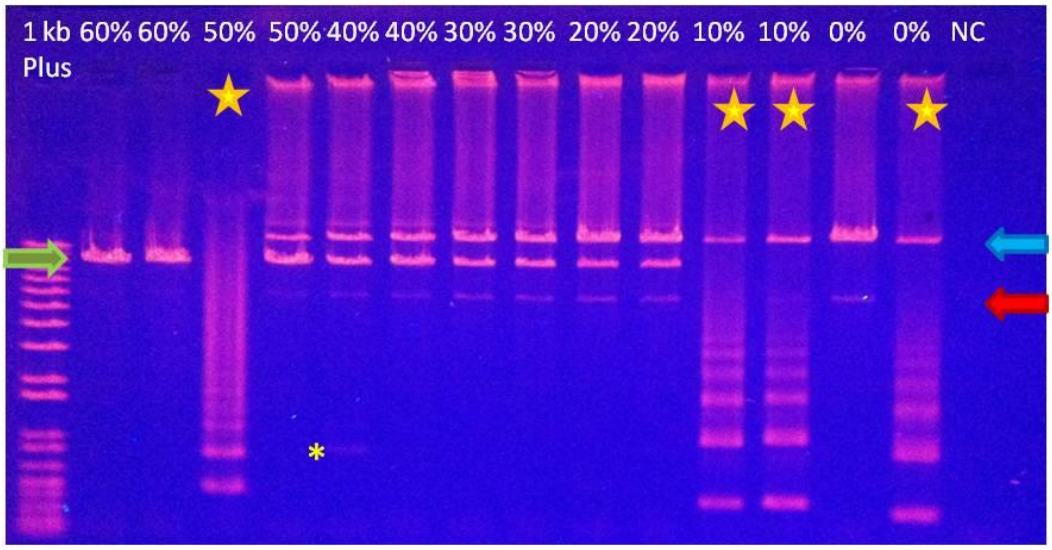


Figure 18. PrimeSTAR® GXL Human MtDNA Mixture Study.

Products visualized on gels stained with Sybr® Gold (A) and ethidium bromide (B). Sybr® Gold is a more sensitive stain and results in better visualization of background bands (red arrow and yellow asterisks). Lane 1 is a 1kb Plus DNA Ladder. 60% is the KSS sample. 0% is the HL60 sample. 50% to 10% are increasing dilutions of the KSS sample mixed with the HL60 sample. The deletion-containing amplicon was readily detected at ratios ranging from 60% to 20%, but was only observed in one of the 10% duplicate reactions (aligned with green arrow). The intact amplicon was readily observed in all reactions in which HL60 DNA had been added (aligned with blue arrow), with the exception of one of the 50% duplicate reactions. Smearing and nondescript banding were observed (yellow stars). NC is a PCR negative control.

In order to further resolve the minimum ratio for detection of the deletion-containing amplicon, the ratio experiment was continued testing ratios of 16.5%, 13.5%, 10%, and 5%. The deletion-containing amplicon was readily observed for all mixed samples (Figure 19, aligned with green arrow). The intact amplicon was also observed for the samples diluted with HL60 DNA (Figure 19, aligned with blue arrow). Present in all reactions was the primary background band (Figure 19, aligned with red arrow) visualized in other amplifications of human DNA (Figures 16 and 17, aligned with red arrow). As with the previous experiment, there were reactions that displayed an unusual amount of smearing and nondescript banding (Figure 19, yellow stars). However, samples were tested in duplicate, and this phenomenon was only observed in one of the duplicate reactions per sample. This is unexpected since all reaction components for duplicate reactions were prepared in a single master mix. Therefore, reactions were identical and these results indicate that the observed phenomena is a stochastic event. In addition, a single reaction displayed a second background band that was not observed of the other amplifications (Figure 19, yellow asterisk). This was unexpected since all reactions contain template DNA from the same two samples (*ie.* KSS and HL60). Both of these phenomena were noted in the previous experiment, though no cause was determined since these affects occurred without any apparent pattern. Regardless of these stochastic effects, this experiment indicates that the assay is able to detect deletion-containing mtDNA even when outnumbered by intact mtGenomes twenty-to-one.

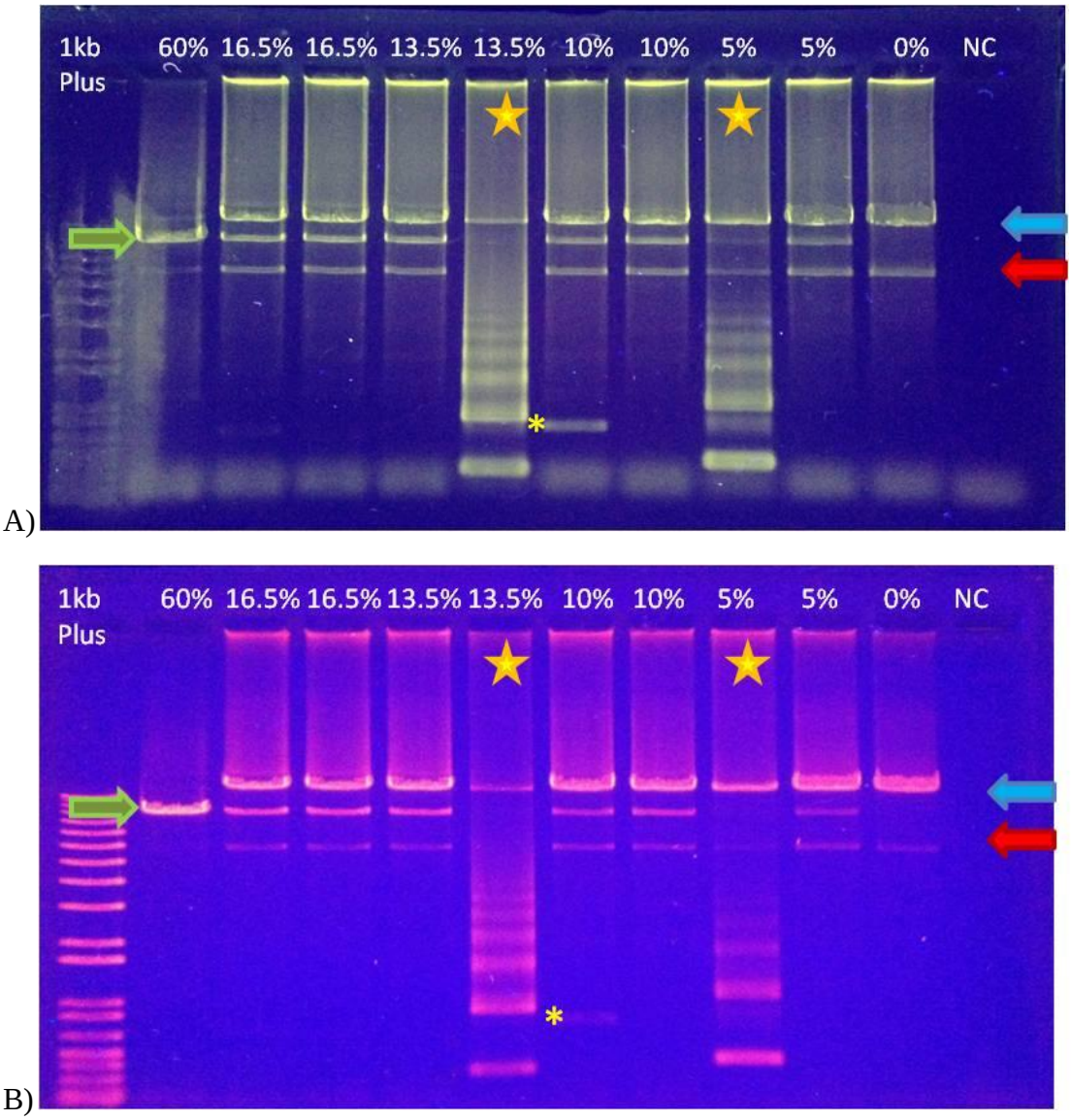


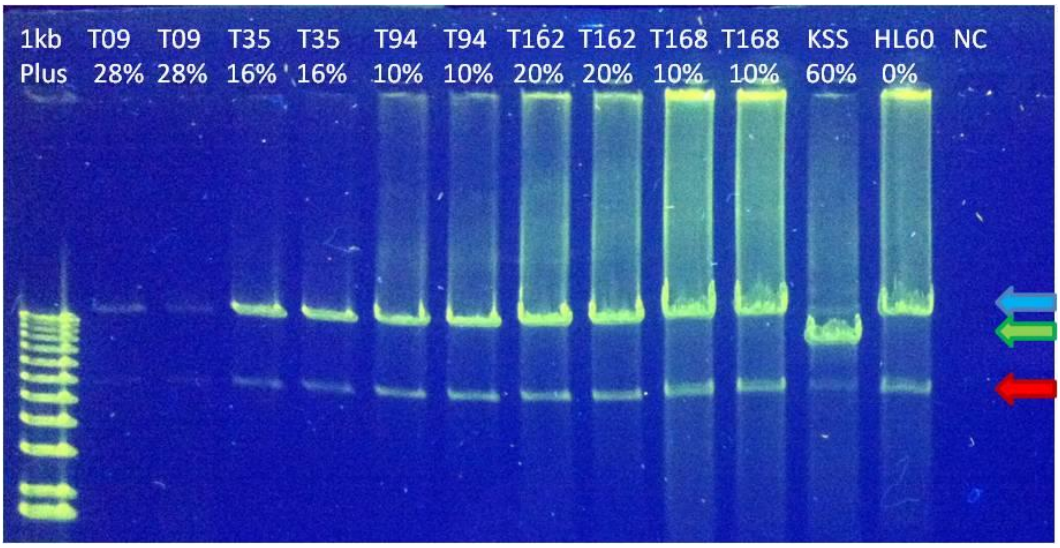
Figure 19. PrimeSTAR® GXL Human MtDNA Mixture Study II.

Products were visualized using Sybr® Gold (A) and ethidium bromide (B). Lane 1 is a 1kb Plus DNA Ladder. 60% is the KSS sample. 0% is the HL60 sample. 16.5% to 5% are increasing dilutions of the KSS sample mixed with the HL60 sample. The intact amplicon was readily observed in all reactions in which HL60 DNA had been added (aligned with blue arrow). The deletion-containing amplicon was observed in all reactions containing KSS DNA (aligned with green arrow), with the exception of one of the 13.5% and 5% duplicate reactions. These reactions exhibited a large amount of smearing and nondescript banding (yellow stars). For a single reaction with 10% deletion-containing template, a second background band was observed (yellow asterisk). NC is a PCR negative control.

The previous experiment demonstrates the ability of the PrimeSTAR® GXL to detect deletion-containing amplicon in mixtures of the KSS/HL60 DNA comprised of only 5% of deletion-containing template. Because all of the TARCC samples were previously determined to contain between 10% and 28% of deletion-containing templates, the next experiment conducted was to test additional TARCC samples for the detection of deletion-containing mtDNA. Five samples were tested, and their estimated proportion of deletion-containing mtDNA species is listed in Table 2. Samples T94, T162, and T168 were normalized and 7.5pg of mtDNA was amplified. Amplifications of samples T09 and T35 contained 2.36pg and 5.70pg mtDNA, respectively. The KSS and HL60 samples were also amplified as controls for the amplification of deletion-containing and intact amplicons, respectively.

As with previous amplifications of human DNA, the single secondary product was observed in all samples (Figure 20, aligned with red arrow). The expected intact and deletion-containing amplicon was observed for the HL60 and KSS samples, respectively (Figure 20, aligned with blue and green arrows). Only the intact amplicon was observed for the TARCC samples (Figure 20, aligned with blue arrow); there was no evidence of deletion-containing amplicons.

A)



B)

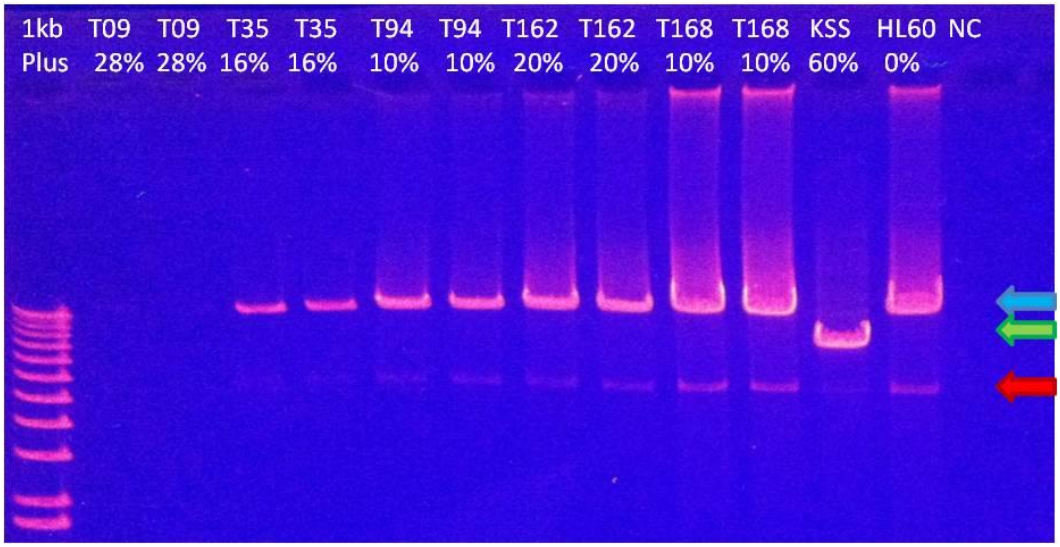


Figure 20. PrimeSTAR® GXL TARCC Sample Amplifications.

Lane 1 is a 1kb Plus DNA Ladder. Samples are labeled along with previously measured proportion of deletion-containing mtGenomes. The expected intact and deletion-containing amplicon was observed for the HL60 and KSS samples, respectively (aligned with blue and green arrows). Only the intact amplicon was observed for the TARCC samples. As with previous amplifications of human DNA, the single secondary product was observed for all (aligned with red arrow). NC is a PCR negative control.

Sanger Sequencing Results

Sanger sequencing of the synthetic mouse mtGenome resulted in 255 bases of sequence (Figure 21). Quality scores ranged from 1 to 55 with an average of 26.6, and 193 (76%) of base positions were assigned a quality score of Q20 or greater. Sixteen bases were assigned ambiguous base calls by the analysis software and were assessed manually (Figure 21, lowercase lettering). The data collection was prematurely terminated by the settings used on the genetic analyzer. Had the run time been set for a longer period, it is likely that data collection would have resulted in more interpretable sequence data. When aligned with the reference genome (NC_005089) using the NCBI Align tool, all 255 bases were complementary to the reference sequence and aligned from positions 14,439 to 14,693.

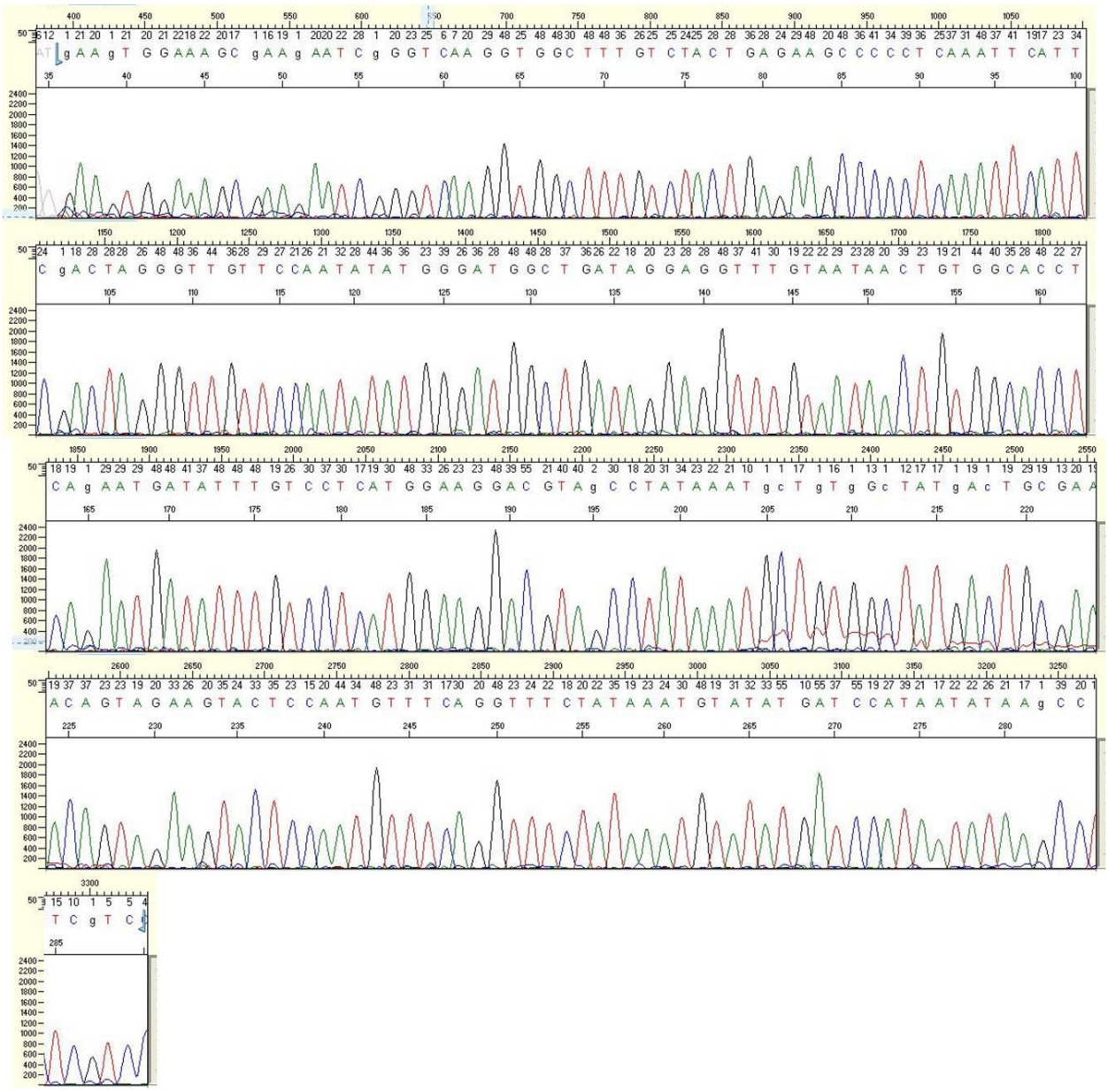


Figure 21. Electropherogram of Sanger Sequencing Data for the Synthetic Mouse MtGenome.

The base calls and base scores are above the peak. Base calls that were manually entered are shown in lowercase lettering.

Next Generation Sequencing Results

Data analysis using NextGENe® software was conducted using the same settings for both the MiSeq System and Ion PGM™ System. Sequencing output files were first converted to FASTA files for both the MiSeq System and Ion PGM™ System. During this conversion step, output files were also filtered based on four quality criteria: 1) median score threshold ≥ 20; 2) maximum number uncalled bases ≤ 3; 3) called base number of each read ≥ 50; and, 4) trim or reject read when ≥3 bases with a score ≤16. Sequences that did not meet these criteria were filtered from the converted FASTA file.

Sequences were then aligned to the reference sequence based on a matching requirement of ≥100 bases and ≥85% homology to the reference. Differences in the sequence data relative to the reference sequence were reported based on whether they were present at >5% of the reads at that position with at least 50 reads. In addition, in order for a difference to be reported, the position must have a minimum of 100× coverage unless the difference was homozygous throughout all reads at that position. A balance ratio was also used in order to eliminate differences that are likely to be false positives. For differences present in ≤80% of the reads at that position, those with a balance of forward-to-reverse alleles of ≤0.200 were removed. The MiSeq System produces paired end reads, and therefore, these data were analyzed with one additional criterion concerning the gap between the paired ends. Paired reads with a gap less than 200bp or greater than 2,000bp were removed from the alignment.

MiSeq System Results

Sequencing of the unamplified synthetic mouse mtGenome resulted in 1,494,600 reads with 140,956 reads (9.43%) removed during the quality filtering. The majority of the reads (99.19%) were removed as a result of a median score of ≤20. Less than 1% of filtered reads were removed because they contained ≥3 bases with a score ≤16 (Table 5).

Table 5. Quality Filtering of Unamplified Synthetic Mouse MtGenome Data from the MiSeq System.

Data are shown for both the forward and reverse sequence reads of the paired end data.

Quality Filter	**Reads Removed**	
Paired Read	Forward	Reverse
Median score threshold ≥ 20	52,523	87,297
Max of uncalled bases ≤ 3	0	0
Called base number of each read ≥ 50	0	0
Trim or reject read when ≥3 bases with a score ≤16	613	523

Of the 1,358,644 reads successfully converted in the FASTA files, 335,845 reads (24.72%) successfully aligned to the reference sequence (NC_005089). The reference sequence was not modified to include the cloning vector so sequence reads representing that portion of the mtGenome were excluded. Sequence reads spanned the entirety of the reference sequence with 99.83% at a minimum of 100× coverage. Coverage at each position ranged from 39 reads to 9,925 reads with an average of 5,017 reads (Figure 22). A total of 99.98% of the positions had base call agreement ≥95%. Coverage and call ratio information is summarized for each nucleotide in Table 6.

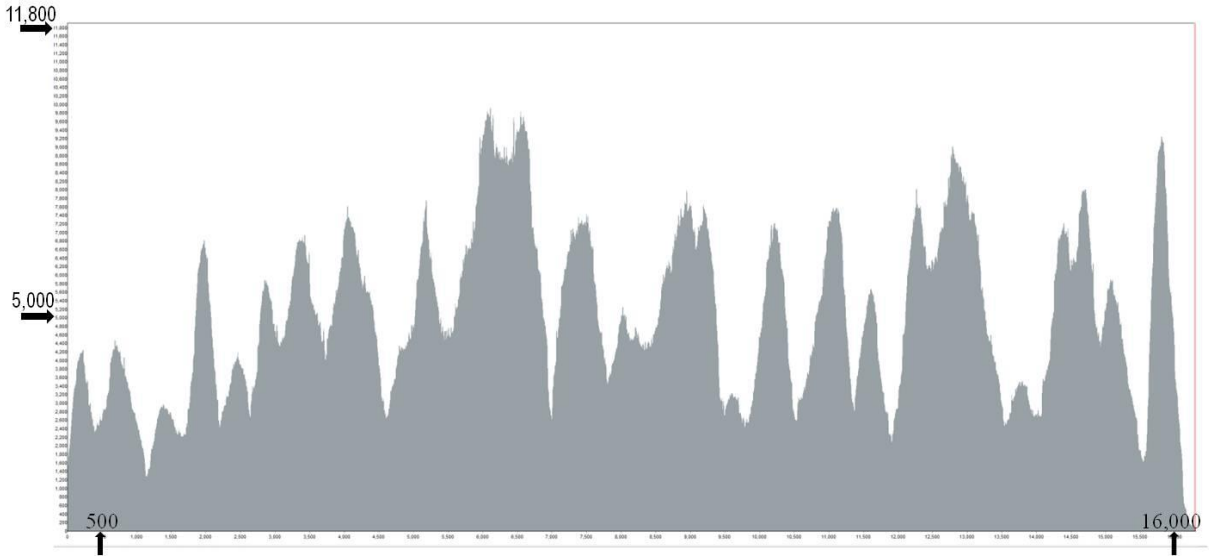


Figure 22. Coverage of the Unamplified Synthetic Mouse MtGenome with the MiSeq System.

Coverage spanned the entire reference genome (NC_005089) with 99.83% at a minimum of 100× coverage. Coverage at each position ranged from 39 reads to 9,925 reads with an average of 5,017 reads. Only a small portion of the mtGenome had less than 100× coverage; it is shown in pink at the far right of the graph (positions 16,273 to 16,298).

Table 6. Sequence Data Summary Information by Nucleotide Type for the Unamplified Synthetic Mouse MtGenome with the MiSeq System

Summary Measure	Adenine	Cytosine	Guanine	Thymine
Total Bases	5,628	3,976	2,013	4,681
Minimum Coverage	39×	41×	83×	42×
Maximum Coverage	9,911×	9,925×	9,826×	9,899×
Average Coverage	4,844×	5,181×	5,243×	4,990×
Total with 100% Call Ratio	7.60%	8.20%	6.61%	5.60%
Average Call Ratio	99.87%	99.91%	99.88%	99.86%
Minimum Call Ratio	94.59%	97.58%	95.63%	91.41%

Only a single difference was reported in the sequence data of the unamplified synthetic mouse mtGenome. At position 5,182, there was a deletion of an adenine. This position had coverage of 7,409 reads and the deletion occurred in 5.29% of the sequence reads at that position. Position 5,182 is the last adenine in a homopolymeric stretch of eleven adenines (Figure 23). This is the only homopolymeric stretch of that length in the *Mus musculus* genome (Table 7).

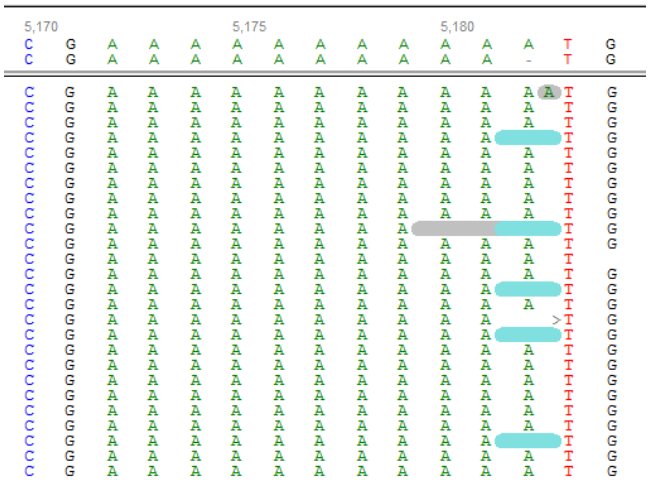


Figure 23. Sequence Alignment of Homopolymeric Region Displaying Deletion of Final Adenine at Position 5182.

A homopolymeric stretch consisting of 11 adenines spans positions 5,171 to 5,182. A deletion of the final adenine at position 5,182 occurred in 5.29% of sequence reads spanning this region.

Table 7. Homopolymeric Regions of the Synthetic Mouse MtGenome.

The homopolymeric region is characterized by the number of nucleotides it contains. The number of homopolymeric regions are tallied for each nucleotide.

Length	Adenine	Cytosine	Guanine	Thymine
Five	33	9	3	10
Six	9	3	0	0
Seven	7	1	0	0
Eight	2	0	0	0
Eleven	1	0	0	0

The sequence of the synthetic mouse mtGenome from positions 16,100 to 16,108 is: 5'-CCCCCTCCT-3'. The positions 16,105 and 16,108 had low base call agreement. These positions had coverage of 1,083 reads and 1,029 reads, respectively. The majority of reads indicated these positions to be thymines, which is concordant with the reference genome. However, the data were composed of cytosines in 5% and 8% of the sequence reads at those positions, respectively. These differences were not reported by the analysis software because cytosines were only observed in sequence data in the forward direction and are likely sequencing errors.

Sequencing of the gel-purified synthetic mouse mtGenome LR PCR amplicon resulted in a total of 272,726 reads with77,138 reads (28.28%) removed during the quality filtering. A total of 99.81% of filtered reads was removed as a result of a median score ≤10. Less than 1% of the filtered reads was removed because they contained ≥3 bases with a score ≤16 (Table 8).

Table 8. Quality Filtering of Gel-Purified Synthetic Mouse LR PCR Amplicon.

Data are shown for both the forward and reverse sequence reads of the paired end data.

Quality Filter	Reads Removed	
Paired Read	Forward	Reverse
Median score threshold ≥ 20	35,237	41,757
Max of uncalled bases ≤ 3	0	0
Called base number of each read ≥ 50	0	0
Trim or reject read when ≥3 bases with a score ≤16	86	58

Of the 195,588 sequencing reads which were successfully converted in the FASTA files, 47,578 reads (24.32%) successfully aligned to the reference sequence (NC_005089) with coverage from position 2,072 through 14,741. The amplicon is defined by the limit of the amplification primer set which had five prime positions of 2,066 and 14,745 for the forward and reverse primers, respectively. For this 12,670bp region, 99.40% of the area had a minimum of 100× coverage. Aligned sequence reads span nearly the entire amplicon with the exception of the first few bases of the amplification primer set. Aligned bases had a minimum coverage depth of 1 read and a maximum of 1,843 reads with an average of 914 reads (Figure 24).

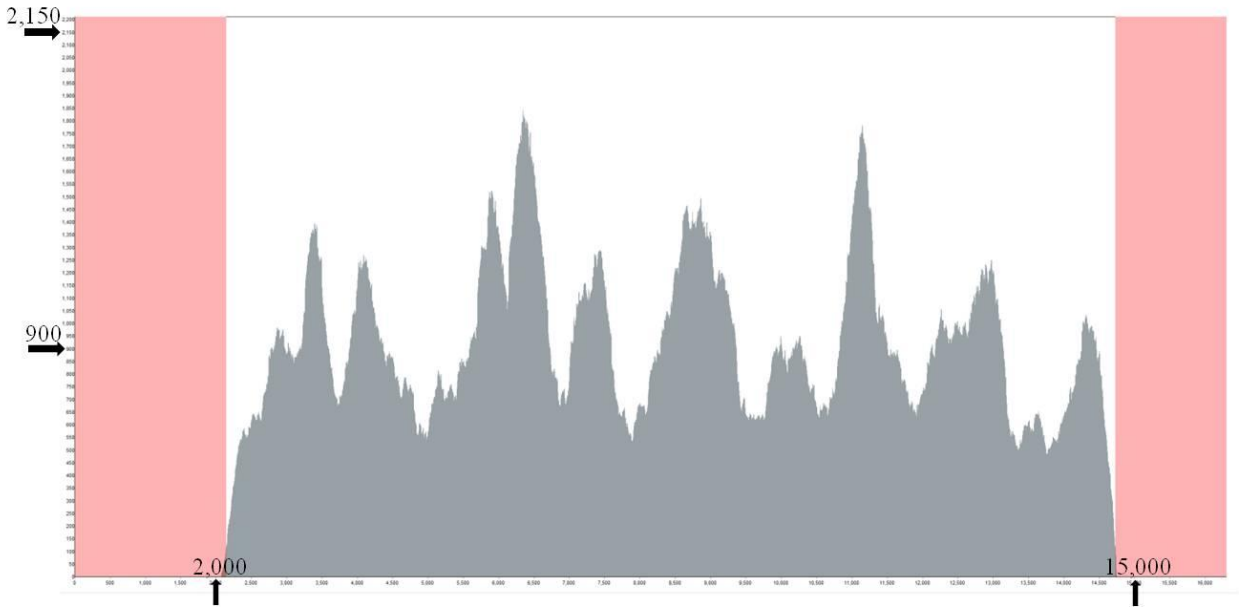


Figure 24. Coverage of the Gel-Purified Synthetic Mouse LR PCR Amplicon with the MiSeq System.

Coverage spanned positions 2072 through 14741 of the reference genome (NC_005089) with 99.3% at a minimum of 100× coverage. Coverage at each position ranged from 1 read to 1,843 reads with an average of 914 reads. Areas with less than 100× coverage are shown in pink.

The average call ratio across all positions was 99.85% (Table 9). There was only a single position with a call ratio less than 95%, and only two differences were reported. The first difference was the deletion of an adenine at position 5,182. This position had coverage of 768 reads, and the deletion was present in 13.41% of reads at that positions. The second reported difference was the insertion of an adenine at position 5,183. This position had coverage of 744 reads, and the insertion was present at 8.87% of the reads at that position. In the reference sequence, position 5,182 is the final adenine at the end of a homopolymeric stretch consisting of 11 consecutive adenine bases.

Table 9. Sequence Data Summary Information by Nucleotide Type for the Gel-Purified Synthetic Mouse Amplicon with the MiSeq System.

Summary Measure	Adenine	Cytosine	Guanine	Thymine
Total Bases	4,295	3,163	1,511	3,701
Minimum Coverage	15×	1×	5×	2×
Maximum Coverage	1,816×	1,843×	1,820×	1,820×
Average Coverage	897×	927×	940×	912×
Total with 100% Call Ratio	38.74%	44.29%	34.21%	37.72%
Average Call Ratio	99.86%	99.89%	99.85%	99.85%
Minimum Call Ratio	86.59%	98.35%	98.73%	98.25%

Sequencing of the background band co-amplified with the HL60 sample resulted in 878,200 sequence reads with 147,326 reads (16.78%) removed during the quality filtering. A total of 99.67% of the sequence reads was removed due to a median score threshold ≤20. Less than 1% of the sequence reads was removed because they contained ≥3 bases with a score ≤16 (Table 10).

Table 10. Quality Filtering of Background Band Co-Amplified with the HL60 Sample During LR PCR.

Quality Filter	Reads Removed	
Paired Read	Forward	Reverse
Median score threshold ≥ 20	63,984	82,849
Max of uncalled bases ≤ 3	0	0
Called base number of each read ≥ 50	0	0
Trim or reject read when ≥3 bases with a score ≤16	286	207

Of the 730,874 sequence reads successfully converted in the FASTA files, 230,462 reads (31.53%) aligned to the rCRS (NC_012920). A total of 73.04% of the rCRS was represented with a minimum of 100× coverage. Coverage depth ranged from 1 read to 15,374 reads with an average of 3,195 reads (Figure 25). A number of differences between the sequence data and the reference genome were reported (Table 11). Sequence data at two positions, 2,445 and 5,149, indicated heteroplasmy at <10%. The full HL60 mtGenome was previously sequenced at UNTHSC using the same platform and analysis tools. The differences observed in the non-specific background band were consistent with those characterized previously in sequence data of the full HL60 mtGenome in the overlapping regions (data not shown). Therefore, these results confirm that the sequence data obtained from the gel purified non-specific background band are a result of amplified mtDNA from the HL60 sample.

Table 11. Differences Reported for the Background Band Coamplified with the HL60 sample with the MiSeq System.

Reference Position	Reference Nucleotide	Coverage (×)	Allele Call	Allele Frequency (%)
263	A	233	A>G	100
295	C	279	C>T	99.64
316	G	343	insC	80.17
489	T	673	T>C	99.55
750	A	6,841	A>G	99.77
1,438	A	9,653	A>G	99.63
2,445	T	9,976	T>CT	6.8
2,706	A	9,858	A>G	99.92
3,107	A	9,301	delA	99.82
4,216	T	8,292	T>C	99.32
4,769	A	9,883	A>G	99.8
5,149	C	11,471	C>CT	6.09
5,228	C	11,515	C>G	99.7
5,633	C	3,595	C>T	99.81
7,028	C	56	C>T	100
8,860	A	56	A>G	100
11,251	A	446	A>G	99.33
11,719	G	946	G>A	99.58
12,071	T	518	T>CT	41.89
12,612	A	580	A>G	99.14
13,708	G	184	G>A	100
14,569	G	586	G>A	99.66
14,766	C	940	C>T	99.57
15,257	G	1,074	G>A	99.81
15,326	A	1,003	A>G	99.6
15,452	C	849	C>A	98.94
15,812	G	551	G>A	99.82
16,069	C	732	C>T	99.45
16,193	C	851	C>T	97.88
16,278	C	780	C>T	98.85
16.362	T	467	T>C	97.86

This band is a result of two non-specific products (Figure 25). Based on an increase in the coverage, it appears as though the first non-specific product (primary amplicon) (Figure 25, green bar) is amplified with greater efficiency than the second (secondary amplicon) (Figure 25, purple bar).

A)

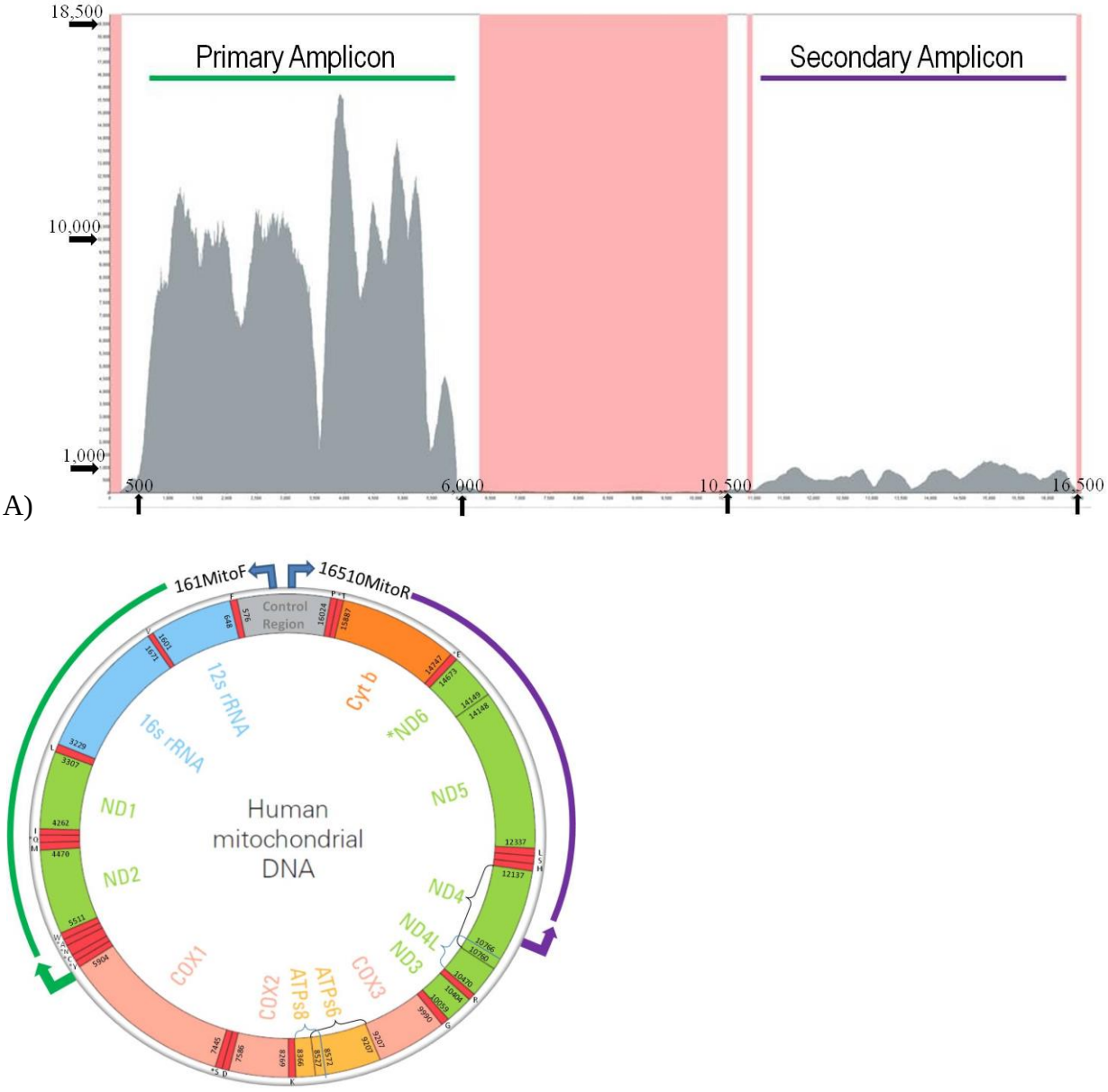


Figure 25. Coverage of the Background Band Co-Amplified with the HL60 Sample During LR PCR.

A) Data were aligned to the rCRS (NC_012920) with 73.04% of the reference represented with a minimum 100× coverage. Areas with less than 100× coverage are shown in pink. The green bar indicates the main non-specific amplicon that spans ~6kb of the mtGenome beginning with the 161MitoF primer and terminating around position 6,000. The purple bar indicates a second non-specific amplicon that is also ~6kb in size, and spans from around position 10,500 to the 16510MitoR primer. B) Cartoon of the human mtGenome indicating regions co-amplified as non-specific products during LR PCR. Green and purple bars correspond to those in (A).

The primary amplicon appears to be the product of non-specific primer binding in the region surrounding reference position 6,000. Coverage in this area has a marked decline from >900 reads at position 5,912 to<300 reads at position 5,913. Coverage then steadily declines until position 6,480 where it remains consistently between 50 and 60 reads. Therefore, non-specific priming to generate this amplicon likely occurs somewhere in this region between positions 5,900 and 6,500.

The NCBI Align Tool was used to evaluate the primer set for homology elsewhere in the human mtGenome using the rCRS (NC_012920). There was only a single site in the human mtGenome with homology to the 3' end of the forward primer. The final seven bases at the 3' end of the 161MitoF forward primer had homology to the complementary strand from position 2,497 through 2,491. This indicates that the 161MitoF primer is capable of priming in the reverse direction from that site. There was also a single site in the human mtGenome with homology to the 3' end of the reverse primer. The final seven bases at the 3' end of the 16510MitoR primer had homology to the human mtGenome from position 4,034 through 4,040. This would indicate that the 16510MitoR primer is potentially capable of priming in the forward direction from this site. However, these are not the locations indicated by the sequence data as the site of priming for either non-specific amplicon observed from the background band.

The NCBI Align tool was used for comparison of the primer sequences to the rCRS just for the region of likely reverse priming (positions 5,900-6,500 of the rCRS) for the primary amplicon. No significant similarity was found between the 16510MitoR primer and this region. The 161MitoF primer has seven bases in a row (bases 17-23) that have homology to the complementary strand from positions 6,092 to 6,098. This homology indicates the potential for the 161MitoF primer for priming in the reverse direction at position 6,088 to 6,114. The

Oligoanalyzer V3.1 (Integrated DNA Technologies) tool was used to evaluate potential base pairing between the entire primer and this region (Figure 26).

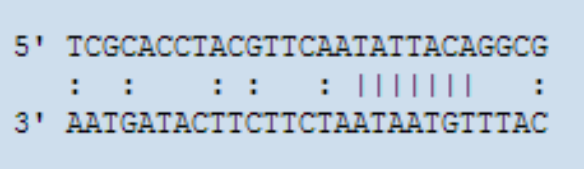


Figure 26. Potential Base Pairing Between 161MitoF and Positions 6,088 to 6,114 of the rCRS.

161MitoF is the top sequence (5' to 3') while the rCRS from position 6,088 to 6,114 is shown 3' to 5'. Strong base pairing is shown with a solid line, while weaker associations are shown with a dotted line.

For the second region of non-specific amplification, coverage hovers around 70 reads for positions 10,408 to 10,440 before beginning to steadily increase. Also, if the main background band is the result of priming by the 161MitoF at position 161 and non-specific priming of 161MitoF at position 6,114, that would result in a 6kb amplicon. For the secondary non-specific product to migrate the same distance in the gel, it would have to be approximately the same size. If this secondary product is the result of amplification by the 16510MitoR primer at position 16510 and a non-specific forward priming site, that non-specific priming site would need to be located near position 10,557 for the product to be the same size as the main non-specific product observed here. Therefore, the NCBI Align tool was used to evaluate positions 10,400 to 10,600 of the rCRS for homology to the primer set. The 16510MitoR primer had no significant homology to this region. The 161MitoF has seven bases in a row (bases 15-21) that have homology to the rCRS from position 10,489 to 10,495. This homology indicates the potential for

forward priming at positions 10,475 to 10,501. The Oligoanalyzer V3.1 tool was used to evaluate potential base pairing between the entire primer and this region (Figure 27).

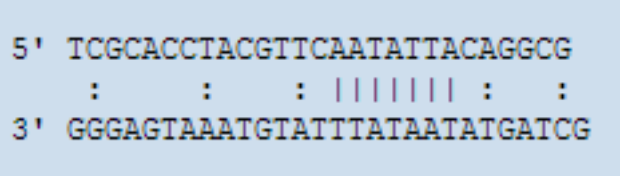


Figure 27. Potential Base Pairing Between 161MitoF and Positions 10,745 to 10,501 of the rCRS

161MitoF is the top sequence while the rCRS from position 10,475 to 10,501 is shown beneath. Strong base pairing is shown with a solid line, while weaker associations are shown with a dotted line.

Ion PGM™ System Results

Sequencing of the unamplified synthetic mouse mtGenome resulted in 79,097 reads with 14,436 reads (18.25%) removed during the quality trimming. More than half of the filtered reads (69.20%) were removed because they contained ≥3 bases with a score ≤16. Additionally, 16.65% and 14.15% of the sequence reads were removed due to a median score ≤20 and a called base number ≤50, respectively.

Table 12. Quality Filtering of the Unamplified Synthetic Mouse MtGenome from the Ion PGM™ System.

Quality Filter	Reads Removed
Median score threshold ≥ 20	2,404
Max of uncalled bases ≤ 3	0
Called base number of each read ≥ 50	2,042
Trim or reject read when ≥3 bases with a score ≤16	9,990

Of the 64,661 reads in the converted FASTA file, 12,379 reads (19.14%) successfully aligned to the reference sequence (NC_005089). Sequence reads spanned the entire reference

genome, and 88.54% of the genome had a minimum of 100× coverage. Coverage ranged from 11 reads to 263 reads with an average of 154 reads (Figure 28). A total of 94.53% of positions had 100% agreement of base calls (Table 13).

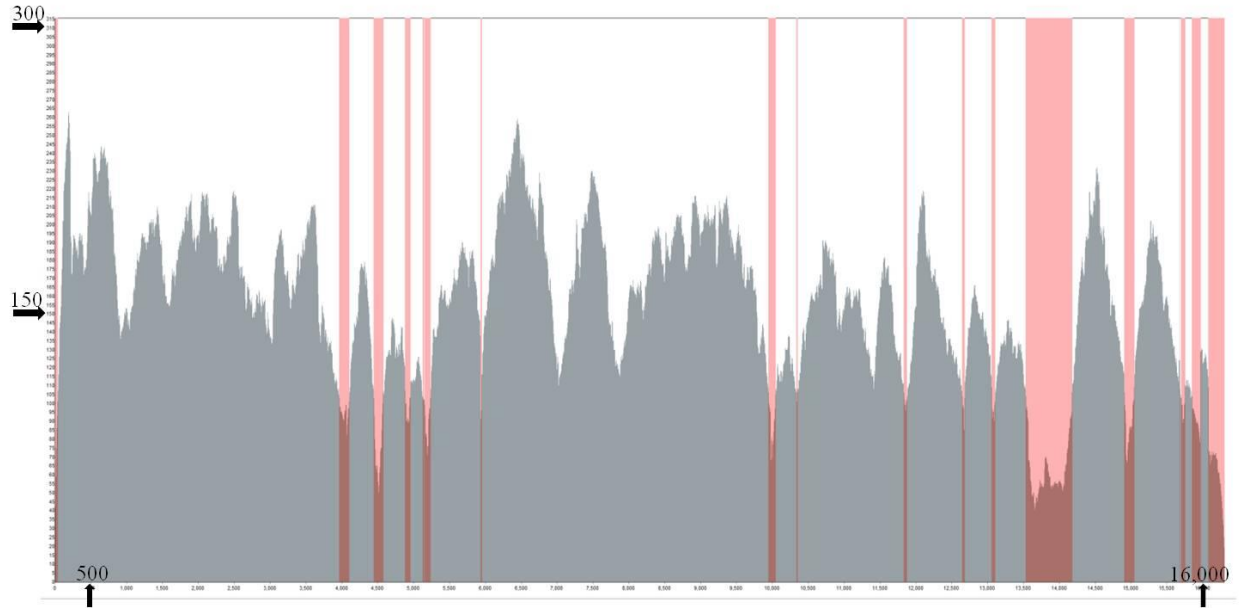


Figure 28. Coverage of the Unamplified Synthetic Mouse MtGenome with the Ion PGM™ System.

Coverage spanned the entire reference genome (NC_005089), and 88.54% of the genome had a minimum of 100× coverage. Coverage ranged from 11 reads to 263 reads with an average of 154 reads. Areas with coverage below 100× are shown in pink.

Table 13. Sequence Data Summary Information by Nucleotide Type for the Unamplified Synthetic Mouse MtGenome with the Ion PGM™ System.

Summary Measure	Adenine	Cytosine	Guanine	Thymine
Total Bases	5,628	3,976	2,013	4,681
Minimum Coverage	11×	11×	25×	11×
Maximum Coverage	261×	263×	261×	258×
Average Coverage	153×	151×	162×	156×
Total with 100% Call Ratio	93.82%	94.24%	95.78%	95.04%
Average Call Ratio	99.79%	99.90%	99.94%	99.92%
Minimum Call Ratio	51.22%	66.67%	91.91%	80.84%

There was only a single difference reported in sequence data for the unamplified synthetic mouse genome. This was the deletion of an adenine at position 1,494. The position had coverage of 171×, and the difference occurred at a frequency of 31.58% of reads at this position. This position is the final adenine in a homopolymeric stretch of seven. In addition, there were 97 positions in which sequence data consisted of deletions at >5% and 153 positions with an insertion ratio of >5%. All of these indel positions were associated with homopolymeric stretches ranging in size from two to 11 bases. However, these positions were not reported as differencess because they either: 1) did not have a minimum read coverage of 50 reads or 2) read coverage was not balanced in the forward and reverse directions to a minimum ratio of 0.200.

Sequencing of the LR PCR amplified synthetic mouse mtGenome resulted in 61,946 reads with 11,713 reads(18.91%) removed during the quality filtering. More than half of the filtered reads (68.57%) were removed because they contained ≥3 bases with a score ≤16. Additionally, 14.33% and 17.10% of the filtered reads were removed due to a median score ≤20 and a called base number ≤50, respectively (Table 14).

Table 14. Quality Filtering of LR PCR Amplified Synthetic Mouse MtDNA from the Ion PGM™ System.

Quality Filter	Reads Removed
Median score threshold ≥ 20	1,678
Max of uncalled bases ≤ 3	0
Called base number of each read ≥ 50	2,003
Trim or reject read when ≥3 bases with a score ≤16	8,032

Of the 50,233 reads successfully converted in the FASTA file, 37,811 reads (75.27%) aligned. Sequence reads spanned 12,680 bases of the reference genome, and 99.81% of that region had a minimum of 100× coverage. Coverage ranged from 74 reads to 833 reads with an average of 554 reads (Figure 29). A total of 82.15% of the positions had 100% agreement of base calls (Table 15).

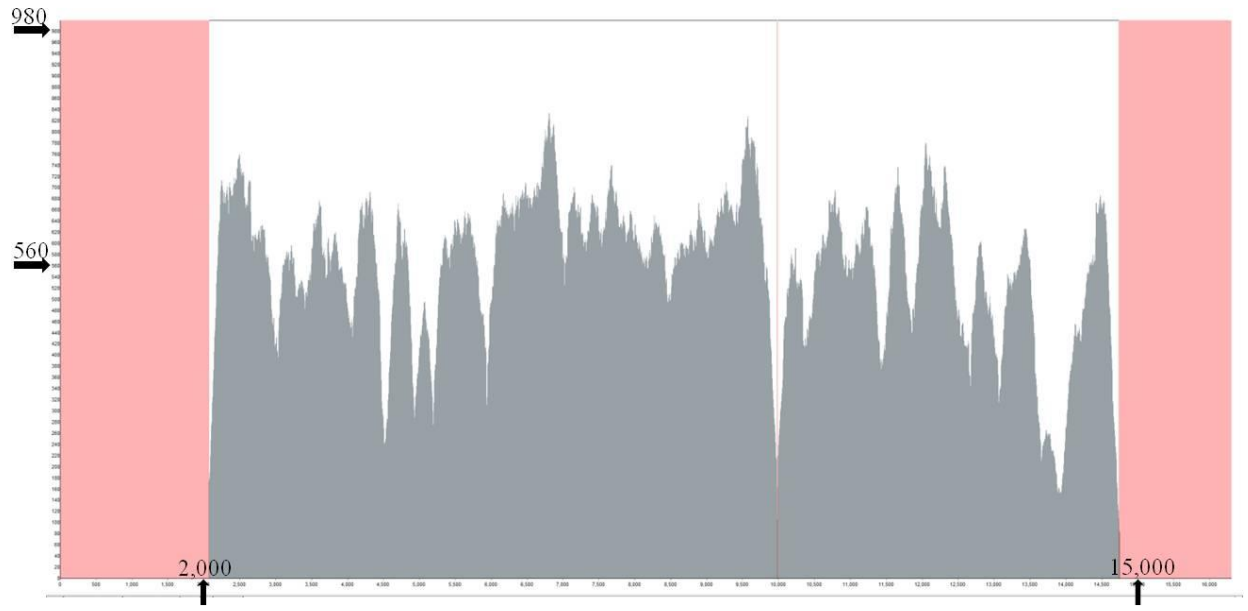


Figure 29. Coverage of the LR PCR Amplified Synthetic Mouse MtDNA with the Ion PGM™ System.

Sequence reads spanned 12,680 bases of the reference genome, and 99.81% of that region had a minimum of 100× coverage. Coverage ranged from 74 reads to 833 reads with an average of 554 reads. Areas with less than 100× coverage are shown in pink.

Table 15. Sequence Data Summary Information by Nucleotide Type for the LR PCR Amplified Synthetic Mouse MtDNA with the Ion PGM™ System.

Summary Measure	Adenine	Cytosine	Guanine	Thymine
Total Bases	4,298	3,167	1,512	3,703
Minimum Coverage	96×	74×	90×	81×
Maximum Coverage	829×	821×	827×	833×
Average Coverage	553×	540×	566×	564×
Total with 100% Call Ratio	84.64%	79.41%	76.79%	83.74%
Average Call Ratio	99.69%	99.88%	99.92%	99.88%
Minimum Call Ratio	38.98%	64.83%	86.12%	75.22%

There were 19 differences reported in the sequence data of the LR PCR amplified synthetic mouse mtGenome. These differences consisted of 2 insertions and 17 deletions, and all were associated with homopolymeric stretches (Table 16). In addition, there were 70 positions with a deletion present at >5% of sequence reads and 32 positions with an insertion ratio of >5%. All of these positions were associated with homopolymeric stretches ranging in size from 2 to 11 bases. However, these positions were not reported because they either: 1) did not have a minimum read coverage of 50 reads or 2) read coverage was not balanced in the forward and reverse directions to a minimum ratio of 0.200.

Table 16. Differences Reported for Sequencing of the LR PCR Amplfied Synthetic Mouse MtDNA with the Ion PGM™ System.

Reference Position	Reference Nucleotide	Coverage (×)	Allele Call	Allele Frequency (%)	Relation to Homopolymeric Stretch
2,276	T	667	insA	12.89	follows 6 adenines
9,829	T	547	insA	11.33	follows 8 adenines
3,625	A	656	delA	13.72	last adenine in a stretch of 6
4,059	A	413	delA	60.29	last adenine in a stretch of 8
4,319	A	668	delA	10.03	last adenine in a stretch of 6
4,731	A	627	delA	15.15	last adenine in a stretch of 6
4,850	A	524	delA	13.74	last adenine in a stretch of 5
5,182	A	243	delA	36.21	last adenine in a stretch of 11
9,828	A	550	delA	30.91	last adenine in a stretch of 8
10,233	A	574	delA	39.72	last adenine in a stretch of 7
10,424	A	459	delA	34.86	last adenine in a stretch of 7
10,445	A	464	delA	11.21	last adenine in a stretch of 5
11,565	A	627	delA	26.48	last adenine in a stretch of 6
12,026	T	739	delT	10.55	last thymine in a stretch of 5
12,642	A	386	delA	41.45	last adenine in a stretch of 7
12,816	A	596	delA	10.23	last adenine in a stretch of 5
12,834	A	567	delA	13.4	last adenine in a stretch of 5
14,037	A	350	delA	21.43	last adenine in a stretch of 5
14,057	A	371	delA	18.6	last adenine in a stretch of 5

CHAPTER 4

DISCUSSION

LR PCR is commonly used to detect mitochondrial deletions associated with diseases. This method produces spurious background bands and was not successful at detecting low level mtDNA deletions. Initial amplification experiments utilizing the SequelPrep™ assay confirmed this method is not as specific as previously reported. Background bands were observed to be numerous and inconsistent in size between replicates of the same sample. Due to the random nature of the background bands observed with this assay, they do not represent true deletion-containing amplicons. These background bands may be the result of intra-molecular strand transfers by the polymerase due to conformation of the circular template in the early cycles of the PCR. It is possible that strand transfer may occur when the supercoiled mtGenome is the dominant template. Then, due to the resulting amplicons being smaller in size than the intact amplicon, these random products may be preferentially amplified leading to multiple, distinct background bands. In order to test this hypothesis, a future direction for this research may be to gel-purify and sequence the background bands. If sequences aligned to multiple, non-linear sites within the mtGenome, that would support the proposed explanation of their origin. Then to determine whether this was the result of circular, supercoiled template, the mtGenome could be linearized prior to amplification. This could be easily accomplished with an appropriate restriction enzyme with specificity to a single site outside of the template region. If that eliminated background bands, then it would be reasonable to conclude that the hypothesis was correct.

Some success in decreasing the number of background bands was achieved. With the successful amplification of the intact amplicon, the overall number of background bands was reduced through increasing of annealing temperature, lowering cycle number, and decreasing the quantity of DNA added to the reaction. Several other factors may have been explored to optimize the SequelPrep™ Assay. For example, the SequelPrep™ Long PCR kit is supplied with two enhancers, A and B, that may be added at varying concentrations. Also, other components such as primer concentration or extension time may have been modified.

Further optimization efforts were directed towards the PrimeSTAR® GXL assay due to better performance in initial experiments. In amplifications of human mtDNA with the PrimeSTAR® GXL assay, the occasional presence of excess smearing and indistinct banding was observed. . This phenomenon appeared randomly, and would often affect only a single reaction of those samples prepared in duplicate. This may be the result of concatamerization by the PrimeSTAR® GXL polymerase. Future experiments could be conducted to confirm this hypothesis.

Interestingly, the intact amplicon was either not observed or only very faintly detected in amplifications of the KSS DNA sample. Given that the proportion of intact mtGenomes was estimated to comprise roughly 40% of the sample, these results were unexpected. Also, with a small addition of HL60 DNA to the KSS DNA in the mixture study, the intact amplicon was easily observed. This is evident in the difference seen between samples where HL60 DNA was added to reduce the deletion-containing mtGenome from 60% (pure KSS DNA) to 50% (Figure 17). This illustrates the concept introduced here that sometimes 'controls' have unexpected characteristics. For instance, it may be possible that the intact amplicon was actually at a lower proportion than expected, which is why it was not readily detected. If this is the case, the ratios

calculated may be in error and actually contain a larger amount of deletion-containing template than reported by the qPCR assay. A future direction of this research may be to reconstruct the ratio experiments using the synthetic mouse mtGenome. The synthetic mtGenome may be easily manipulated using restriction digests and ligation to generate deletion-containing molecules in which the deletion is of a precise size based on enzyme specificity. Mixtures of intact and deletion-containing synthetic mtGenomes may be less likely to exhibit the unanticipated result demonstrated by the KSS sample. This may result in a more accurement measurement of the limit of deletion detection for this assay.

Failure to detect deletions in the TARCC samples was likely due to the nature of the deletions present in these samples. The assay to estimate deletion content can only indicate a relative proportion of deletion-containing mtGenomes to intact mtGenomes. It is incapable of characterizing those deletions in regard to size or the number of different types of deletions present. Given that the DNA for these samples was isolated from peripheral blood where cell turnover is high, it is highly plausible that the samples do not contain a single deletion type, but rather many types of deletions all present at varying levels. For example, sample T168 had an estimated deletion proportion of 10% by qPCR (Table 2). If there were 10 different types of deletions present all varying in size and number, it is possible that all of the deletion-containing mtGenomes were present at a proportion too minimal to be detected with the LR PCR assay. A possible way to address this in future experiments using the synthetic mouse mtGenome would be to use specific restriction enzymes to generate deletion-containing molecules of different sizes. These altered mtGenomes could be used to perform a more elaborate ratio experiment to more accurately represent the dynamics of intact and deletion-containing mtGenomes expected to be present in blood.

Sequencing of the background band co-amplified with the HL60 sample revealed that it was the result of two separate amplicons both produced by non-specific priming by the 161MitoF forward primer. This highlights the importance of primer design in LR PCR applications. The 161MitoF/16510MitoR primer set was on the lower end of the oligonucleotide length (25-35mer) and Tm (≥65°C) recommended in the PrimeSTAR® GXL manual. Attempts were made to alter the primers by increasing their length and Tm. However, *in silico* analysis of these alterations demonstrated increased homology to areas elsewhere in the mtGenome. A future direction of this research could be to implement the LR PCR protocol with alternative primer sets to eliminate these non-specific products.

Sequence data were successfully obtained for the synthetic mouse mtGenome using Sanger sequencing and both next generation platforms, the MiSeq System and Ion PGM™ System. For Sanger sequencing, sequence information was obtained for 255 bases; additional sequence information would likely have been obtained with a longer run time. Baseline noise is apparent throughout the electropherogram (Figure 21). This illustrates why it is difficult to distinguish low levels of heteroplasmy from background noise using this method. This method is also very labor intensive relative to the amount of information obtained in comparison with next generation sequencing methods.

Prior to alignment, sequence data obtained for both the MiSeq System and Ion PGM™ System were filtered for quality based on four criteria. The majority (>99%) of sequence reads removed by quality filtering for the MiSeq System data was filtered out due to a median score threshold ≤20. This indicates that sequence reads containing low quality base calls consisted of low quality base calls throughout the entire read. This differed from what was observed with the Ion PGM™ System. For this system, less than 20% of the reads were removed due to a median

score threshold ≤20. More than half (>65%) of the reads were removed because they contained ≥3 bases with a score ≤16. These results indicate that for the Ion PGM™ system, sequence reads that did not pass quality filtering criteria had low quality base calls mixed with base calls of high quality score. Additionally, between 10 and 15% of reads were discarded because they were less than 50 bases long, while none of the MiSeq System reads were filtered based on this criterion.

During alignment, sequence data were assessed for differences relative to the *Mus musculus* reference mtGenome. The primary criterion was that differences be present at >5% and be represented by at least 50 reads. The majority of the MiSeq System data had base call ratios >95%, and the only differences reported were indels associated with a homopolymeric stretch of 11 adenines. Due to this association, it is much more likely that the indels reported were sequencing errors rather than mutations of the mtGenome.

The Ion PGM™ System produced many more of these sequencing errors. There was only a single reported difference in the sequence data for the unamplified synthetic mouse mtGenome. However, there were an additional 250 indels present at >5% that did not meet reporting criteria. A similar trend was observed in sequence data for the LR PCR amplified synthetic mouse mtGenome with 19 reported differences and an additional 102 indels that were not reported. All of these indels were associated with homopolymeric regions that ranged in length from two to 11 residues. This difficulty in sequencing homopolymeric stretches is consistent with what has been previously observed with this system, and great care needs to be taken in assessing reported differences associated with these motifs to avoid false positive mutations.

A future direction of this research would be to further evaluate all homopolymeric stretches within the sequence data of the synthetic mouse mtGenome to further characterize the

ability of the MiSeq and Ion PGM™ systems to sequence through these regions. Because the true length of the homopolymeric stretches within the *Mus musculus* mtGenome is known, analysis of sequence data from these regions can characterize the error probability for motifs of this type.

In general, the quality filtering and alignment criteria selected were very effective at removing low quality data and sequencing errors. All reported differences using these criteria were associated with homopolymeric stretches, and are likely due to sequencing errors. This presents an opportunity to improve the sequence analysis software to differentiate sequencing errors from true mutations. The sequencing data set of the synthetic mouse mtGenome could be used as a resource for calibration and testing of analysis settings. Future directions may be to employ this data as a practice set in order to assess analysis parameters. Depending on the data that pass quality filtering and are reported during analysis, the user may optimize and validate the software settings for the evaluation of other research samples.

APPENDIX

Induction of High Copy Number of Plasmid within Recombinant *E.coli*

Purpose: In order to harvest plasmid from recombinant cells, cells must first be cultured and induced to produce a high copy number of plasmids within the cells. This is accomplished by culturing cells in LB broth with12.5µg/mL chloramphenicol and treating with CopyControl™ Induction Solution.

Equipment and Supplies

- 500mL sterile Erlenmeyer flask
- Inoculation loop
- 37°C incubator with shaker

Reagents

- LB broth
- 25mg/mL chloramphenicol stock
- CopyControl™ Induction Solution (Epicentre® Biotechnologies)
- PerfectPrep EndoFree Plasmid Maxi Kit (5PRIME)

Procedure

1. Remove frozen stock of *E.coli* culture from -80°C freezer and place on ice along with 25mg/mL chloramphenicol stock from -20°C freezer.
2. Sterilize work area with bleach and light the Bunsen burner.
3. Using aseptic technique, transfer 200mL of LB broth to 500mL Erlenmeyer flask.
   NOTE: Remain close to flame.
   NOTE: Flame top of LB container when opening and before closing.
4. Add 100µL chloramphenicol.
5. Sterilize inoculation loop in flame. Allow inoculation loop to cool. Remaining very close to flame, insert inoculation loop into 1.5mL tube of frozen stock. Transfer to 200mL LB broth + chloramphenicol. Flame top of flask, and cover with foil.
6. Incubate overnight at 37°C without shaking.
7. Remove CopyControl™ Induction Solution and chloramphenicol from -20°C freezer and thaw on ice.
8. Sterilize work area with bleach and light Bunsen burner.
9. Using aseptic technique, remove 160mL of overnight culture. Add 160mL of fresh LB broth and 80µL chloramphenicol.
10. Incubate at 37°C while shaking for 30 minutes.
11. Using aseptic technique, add 200µL of CopyControl™ Induction Solution.
12. Incubate at 37°C with vigorous shaking for 2 hrs.
13. Harvest plasmid per manufacturer's instructions in PerfectPrep EndoFree Plasmid Maxi Kit.

REFERENCES

1. Hatefi Y. The mitochondrial electron transport and oxidative phosphorylation system. Annu Rev Biochem. 1985;54:1015-69.

2. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981 Apr 9;290(5806):457-65.

3. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA. Nat Genet. 1999 Oct;23(2):147.

4. Robin ED, Wong R. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol. 1988 Sep;136(3):507-13.

5. Orrenius S. Reactive oxygen species in mitochondria-mediated cell death. Drug Metab Rev. 2007;39(2-3):443-55.

6. Richter C, Park JW, Ames BN. Normal oxidative damage to mitochondrial and nuclear DNA is extensive. Proc Natl Acad Sci U S A. 1988 Sep;85(17):6465-7.

7. Lynch M, Koskella B, Schaack S. Mutation pressure and the evolution of organelle genomic architecture. Science. 2006 Mar 24;311(5768):1727-30.

8. Pitceathly RD, Rahman S, Hanna MG. Single deletions in mitochondrial DNA--molecular mechanisms and disease phenotypes in clinical practice. Neuromuscul Disord. 2012 Jul;22(7):577-86.

9. Tuppen HA, Blakely EL, Turnbull DM, Taylor RW. Mitochondrial DNA mutations and human disease. Biochim Biophys Acta. 2010 Feb;1797(2):113-28.

10. Monnat RJ,Jr, Loeb LA. Nucleotide sequence preservation of human mitochondrial DNA. Proc Natl Acad Sci U S A. 1985 May;82(9):2895-9.

11. He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. Nature. 2010 Mar 25;464(7288):610-4.

12. Holt IJ, Harding AE, Morgan-Hughes JA. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. Nature. 1988 Feb 25;331(6158):717-9.

13. Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AM, et al. Mitochondrial DNA mutation associated with leber's hereditary optic neuropathy. Science. 1988 Dec 9;242(4884):1427-30.

14. http://www.mitomap.org/MITOMAP

15. Khrapko K, Bodyak N, Thilly WG, van Orsouw NJ, Zhang X, Coller HA, et al. Cell-by-cell scanning of whole mitochondrial genomes in aged human heart reveals a significant fraction of myocytes with clonally expanded deletions. Nucleic Acids Res. 1999 Jun 1;27(11):2434-41.

16. Wilson MR, DiZinno JA, Polanskey D, Replogle J, Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. Int J Legal Med. 1995;108(2):68-74.

17. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977 Dec;74(12):5463-7.

18. Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, et al. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. Science. 1987 Oct 16;238(4825):336-41.

19. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, et al. Fluorescence detection in automated DNA sequence analysis. Nature. 1986 Jun 12-18;321(6071):674-9.

20. Ruiz-Martinez MC, Berka J, Belenkii A, Foret F, Miller AW, Karger BL. DNA sequencing by capillary electrophoresis with replaceable linear polyacrylamide and laser-induced fluorescence detection. Anal Chem. 1993 Oct 15;65(20):2851-8.

21. Swerdlow H, Wu SL, Harke H, Dovichi NJ. Capillary gel electrophoresis for DNA sequencing. laser-induced fluorescence detection with the sheath flow cuvette. J Chromatogr. 1990 Sep 7;516(1):61-7.

22. Karger BL, Guttman A. DNA sequencing by CE. Electrophoresis. 2009 Jun;30 Suppl 1:S196-202.

23. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. accuracy assessment. Genome Res. 1998 Mar;8(3):175-85.

24. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. error probabilities. Genome Res. 1998 Mar;8(3):186-94.

25. Kleparnik K, Foret F, Berka J, Goetzinger W, Miller AW, Karger BL. The use of elevated column temperature to extend DNA sequencing read lengths in capillary electrophoresis with replaceable polymer matrices. Electrophoresis. 1996 Dec;17(12):1860-6.

26. Carrilho E, Ruiz-Martinez MC, Berka J, Smirnov I, Goetzinger W, Miller AW, et al. Rapid DNA sequencing of more than 1000 bases per run by capillary electrophoresis using replaceable linear polyacrylamide solutions. Anal Chem. 1996 Oct 1;68(19):3305-13.

27. Detwiler MM, Hamp TJ, Kazim AL. DNA sequencing using the liquid polymer POP-7 on an ABI PRISM 3100 genetic analyzer. BioTechniques. 2004 Jun;36(6):932-3.

28. Hawes JW, Knudtson KL, Escobar H, Grills GS, Hunter TC, Jackson-Machelski E, et al. Evaluation of methods for sequence analysis of highly repetitive DNA templates. J Biomol Tech. 2006 Apr;17(2):138-44.

29. Hancock DK, Tully LA, Levin BC. A standard reference material to determine the sensitivity of techniques for detecting low-frequency mutations, SNPs, and heteroplasmies in mitochondrial DNA. Genomics. 2005 Oct;86(4):446-61.

30. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. Journal of Biomedicine and Biotechnology. 2012 Apr;e251364.

31. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012 Jul; 24(13):341,2164-13-341.

32. Tonge DP, Pashley CH, Gant TW. Amplicon -based metagenomic analysis of mixed fungal samples using proton release amplicon sequencing. PLoS One. 2014 Apr 11;9(4):e93849.

33. Irwin JA, Saunier JL, Niederstatter 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. J Mol Evol. 2009 May;68(5):516-27.

34. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008 Nov 6;456(7218):53-9.

35. Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008;9:387-402.

36. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011 Jul 20;475(7356):348-52.

37. 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. 2010 Aug 20;5(8):e12295.

38. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012 May;30(5):434-9.

39. Bandelt HJ, Salas A. Current next generation sequencing technology may not meet forensic standards. Forensic Sci Int Genet. 2012 Jan;6(1):143-5.

40. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009 May;6(5):343-5.

41. https://www.neb.com/tools-and-resources/feature-articles/gibson-assembly-building-a-synthetic-biology-toolset.

42. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. Science. 2010 Jul 2;329(5987):52-6.

43. Gibson DG, Smith HO, Hutchison CA, Venter JC, Merryman C. Chemical synthesis of the mouse mitochondrial genome. Nat Methods. 2010 Nov;7(11):901-3.

44. Moraes CT, Kenyon L, Hao H. Mechanisms of human mitochondrial DNA maintenance: The determining role of primary sequence and length over function. Mol Biol Cell. 1999 Oct;10(10):3345-56.

45. Phillips NR, Sprouse ML, Roby RK. Simultaneous quantification of mitochondrial DNA copy number and deletion ratio: A multiplex real-time PCR assay. Sci Rep. 2014 Jan 27;4:3887.

46. Wilson MR, Polanskey D, Butler J, DiZinno JA, Replogle J, Budowle B. Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts. BioTechniques. 1995 Apr;18(4):662-9.

47. Kumar M, Tanwar M, Saxena R, Sharma P, Dada R. Identification of novel mitochondrial mutations in Leber's hereditary optic neuropathy. Mol Vis. 2010 Apr 30;16:782-92.