Oatts, Sarah M., Development of a full mitochondrial genome specific target-enriched library for next generation sequencing applications. Master of Science (Forensic Genetics), May 2013, 65 pp., 4 tables, 13 figures, 27 references.

Mitochondrial DNA (mtDNA) sequencing is used in many applications of the scientific field including disease studies, migration/ancestry studies and forensic identification. Most mitochondrial sequencing thus far has been performed solely on the highly variable control region of the mitochondrial genome, but new technologies such as next generation sequencing platforms are facilitating an increase in whole mitochondrial genome sequencing. In migration studies, obtaining sequences of the entire mtDNA allows for utilization of the more conserved regions of the genome to develop haplotypes which can help in more accurate estimation of population affiliations of specific mtDNA haplotypes.

In this study, a target-enriched full mitochondrial genome library was developed to be used for next generation sequencing technologies. The two 12-plex polymerase chain reaction (PCR) primer multiplexes used to develop this library were designed to be able to completely amplify the mtDNA of a set of Easter Island samples on Whatman™ FTA™ cards that a previously-developed nine primer multiplex reaction could not. It was hypothesized that the matrix of the cards prevented amplification of very long fragments (greater than approximately1000 bp). Successful amplification using the 12-plex reactions was accomplished using extracted DNA samples and direct amplification from the FTA™ cards. Libraries of a pilot set of Easter Island samples were created so that sequences could be generated and genetic associations between the remote island population and its two closest geographic populations, Oceania and mainland Chile, could be determined. The multiplexes developed from this study could be used in the

future for other samples that appear to be challenged due to either sample quality or their complex storage medium, in this case, FTA™ cards.

DEVELOPMENT OF A FULL MITOCHONDRIAL GENOME SPECIFIC TARGET-ENRICHED LIBRARY FOR NEXT GENERATION SEQUENCING APPLICATIONS

THESIS

Presented to the Graduate Council of the

Graduate School of Biomedical Sciences

University of North Texas

Health Science Center at Fort Worth

In Partial Fulfillment of the Requirements

For the Degree of

MASTERS OF SCIENCE

By

Sarah M. Oatts, B.S.

Fort Worth, Texas

May 2013

ACKNOWLEDGEMENTS

I would first like to thank my major professor Dr. John Planz for his tremendous support and guidance throughout this entire process and for challenging me to make this project bigger and better than I ever thought it could be. Special thanks goes to Dr. Michael Allen and David Visi for generously giving me access to the IonTorrent™ and for providing their much needed advice and experience on sequencing my samples. I also must recognize my lab mates Sandy Kirkpatrick and Shantanu Shewale for their assistance in making this project possible. Finally, I would like to thank my husband Matthew, family and friends for their unwavering encouragement and belief in me throughout my Masters program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Mitochondrial DNA sequencing has many applications in the scientific community. The mitochondrial genome is commonly used to determine specific haplotypes of individuals in order to gain a better understanding of the genetics and migration patterns of populations because mtDNA is maternally inherited and has virtually no recombination between generations, allowing haplotypes to be tracked across generations (1, 2). Because of the mitochondrial genome's higher mutation rate, disease studies often look for associations of specific mutations in the mitochondrial sequence within patients and non-effected controls (3). Additionally, mitochondrial DNA analysis is used in forensic cases when samples are thought to be degraded because the mitochondrial genome is more abundant in cells (greater than 1,000 copies per cell versus two nuclear DNA copies per cell), and its circular structure and location within the mitochondrial matrix makes it less prone to degradation than nuclear DNA which is linear (2).

Most mitochondrial haplogroups used in migration studies have been defined by single nucleotide polymorphisms (SNP) in the coding regions or sometimes differences in the sequence of a small region of the mitochondrial genome called the control region. Advancements in sequencing technology, however, have allowed for sequencing of the whole mitochondrial genome (1). By comparing whole mitochondrial genomes instead of using only the control regions or single nucleotide polymorphism (SNP) panels, more differences between individuals can be observed, allowing for greater haplotype resolutions and, therefore, more accurate designations of the ancestry of individuals. Additionally, the hypervariable nature of the control

region is not as useful as the more conserved coding regions of the rest of the mitochondrial genome for tracking migration patterns and creating phylogenetic trees.

One issue in conducting analysis of whole mitochondrial genome sequences is the presence of nuclear copies of mitochondrial DNA (NUMTs). These are portions of the mitochondrial genome that are translocated from the cytoplasmic mitochondria to the nucleus and then inserted into the nuclear genome (3). Although NUMTs have a similar sequence to their mitochondrial counterparts, differences in the mutation rates of mitochondrial DNA versus nuclear DNA can result in some variances in sequence (4, 5). If primers selected for amplification of mitochondrial DNA fragments are homologous to NUMTs, co-amplification can occur, making interpretation and analysis of the targeted sequence difficult (6). Because there are fewer copies of the nuclear genome than the mitochondrial genome, sequence differences due to NUMTs can be mistaken for true low-level mutations in the mitochondrial genome and appear as heteroplasmy (6). The presence of NUMTs in a derived mitochondrial sequence could lead to the misidentification of disease-mutation associations or the erroneous classification of novel sub-haplogroups in phylogenetic studies (7).

The original purpose of this study was to use an already-established multiplex of nine overlapping PCR primer pairs validated by Ramos *et al.* 2009 to amplify the full mitochondrial genomes of a set of samples from Easter Island and then sequence them using a next generation sequencing (NGS) platform. Once, the whole mitochondrial sequences were generated from the Easter Island samples using the Ramos *et al.* primers, they would be compared to previously-published full mitochondrial sequences of individuals from Oceania and mainland Chile, the two populations closest in distance to the remote island, to determine genetic relationships and migration patterns between the three populations. Using whole mitochondrial sequencing for the

comparison of Easter Island individuals and their neighboring populations would result in a better understanding of these genetic associations than other genetic markers have shown.

Several population genetics studies have been conducted to determine the initial colonizers of Easter Island. Lie *et al.* 2007 used mitochondrial DNA (mtDNA) haplotypes, Y-chromosomal short tandem repeats (Y-STRs), single nucleotide polymorphisms (SNPs), and human leukocyte antigen (HLA) haplotypes to compare an Easter Island population to Polynesian, European, and Amerindian populations (8). As most of the individuals used in the study were related, the results could not be applied as an accurate representation of the Easter Island population and statistical analysis to determine genetic distances between populations was not attempted. Instead, the authors only commented on the presence of specific haplotypes in the Easter Island samples that probably originated from Polynesian, European, or Amerindian populations. It was observed that all the sampled mtDNA haplotypes had a probable Polynesian origin. Nine Polynesian and two European Y-STR haplotypes were observed in the sample population. The HLA haplotypes from the population consisted mostly of Polynesian-origin haplotypes but also contained one European and two Amerindian haplotypes. One of the Amerindian HLA alleles observed in the Easter Island samples (B*3905) was found to be frequent in native populations from Paraguay, Bolivia, Brazil, Colombia, Venezuela, and Mexico (8). From the results of this study, it was concluded that the population of Easter Island is most closely associated genetically with Polynesian populations. A subsequent study displayed two more Amerindian HLA haplotypes present in the population, further indication of some Amerindian genetic contribution to the Easter Island population as well (9). The second study did not specify the origin of the newly observed Amerindian HLA haplotypes.

Other studies determined genetic distances based on observed polymorphisms to compare Easter Island natives with other populations. Ghiani *et al.* 2006 compared seven-locus Y-STR haplotypes of Easter Islanders to Pacific, Asian, South American and German populations (10). The South American subpopulations used in the study were not well defined by the authors (Surinamese, native South American and Portenos populations), so it is unknown if native Chilean populations were included in the comparison. The authors determined that the Easter Island population was most closely associated with the Pacific populations, specifically Western Samoans. *Alu* insertion polymorphisms were analyzed by Gonzalez-Perez *et al.* 2006 to evaluate the genetic distance of the Easter Island population to 23 other populations from around the world, including 12 Pacific and four South American populations (11). None of the South American populations used in the comparison were located near mainland Chile. The four South American populations were all from locations on the eastern side of the Andes Mountains, which act as a genetic barrier between Chile and the eastern part of the continent. The shortest genetic distance observed in the study was between the Easter Island population and Southeast Asian island populations, followed by the genetic distance between Easter Islanders and Native Americans.

Although both of these studies included some South American populations in their comparisons, neither incorporated mainland Chileans into their analyses even though the Chilean population is the physically closest of the South American populations to Easter Island. The present study incorporates mainland Chilean population samples from Native American Mapuche individuals as well as previously-published sequences from Oceanians to evaluate genetic relationships to Easter Island individuals. Using known native Chileans for comparison

may provide a more accurate assessment of Easter Island populations to other possible genetic contributor populations.

The original mtDNA specific primers, some created by Ramos *et al.*, and others first described by Torroni *et al.*, amplify fragments approximately 1,500-3,000 bp in length (1, 12) (Figure 1). Ramos *et al.* 2009 performed tests to determine that the primers did not amplify NUMTs (1). The Basic Local Alignment Search Tool (BLAST) available at the National Center for Biotechnology Information was used to ensure the primer sequences were not homologous to any nuclear genome sequences that might lead to amplification of NUMTs (1). Experiments using the primers to amplify both total DNA samples and non-mitochondrial DNA samples also indicated that these primers were specific for the mitochondrial genome (1). The primer set was developed into a single multiplex reaction to facilitate higher throughput databasing.

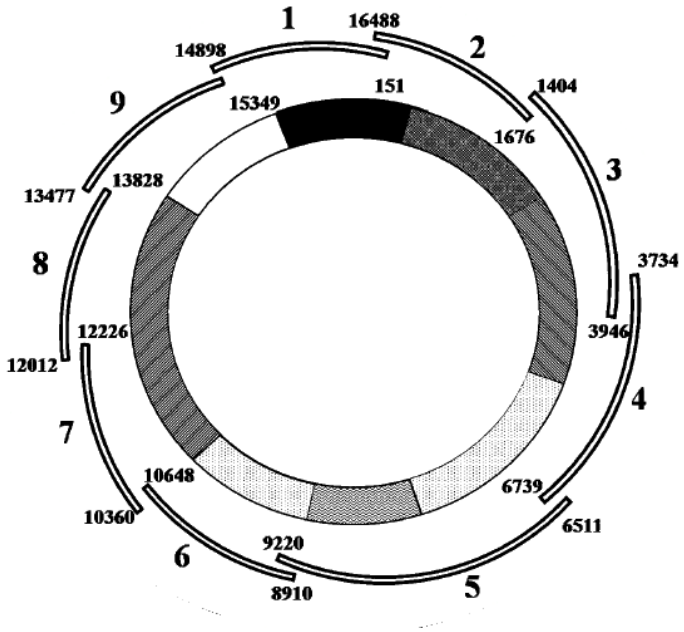


Figure 1: Illustration of the mitochondrial genome and the positions of the nine primer pairs that were going to be used to amplify the entire mitochondrial genome of the Easter Island samples (adapted from Ramos *et al.* 2009).

It was the intention of this study to use this multiplex reaction to develop a targeted library for sequencing a set of Easter Island samples. However initial evaluation of the samples with the multiplex gave poor results. Due to the large sizes of the amplicons it was hypothesized that the storage matrix of the samples (FTA^{TM} cards) prevented the amplification of the fragments.

The revised focus of this study was to develop a multiplex of overlapping primers that will amplify the entire mitochondrial genome in smaller fragments than that of the Ramos *et al.* primers and take a test set of Easter Island and Mapuche natives through the deep sequencing procedure. Based on the preliminary tests, it was hypothesized that by targeting smaller amplicons, the full mitochondrial genomes of the Easter Island database samples could be amplified, allowing for downstream next generation sequencing and comparison of these sequences to those of the mainland Chilean samples and the previously-published Oceanians.

An additional goal of this study was to begin to answer a specific question about the genetic origins of the natives of Easter Island: Are the Easter Islanders more closely related to mainland Chileans or to individuals from Oceania? Based on research on the history of the island and results of previous, related studies (8-11), it was hypothesized that the mitochondrial haplotypes of the people of Easter Island will be more closely associated with those of the populations from Oceania. However an alternate hypothesis, that the Easter Islanders are genetically more similar to Native Chilean populations is equally plausible.

CHAPTER II

MATERIALS AND METHODS

Specimens Examined:

The Easter Island fingerstick blood samples (CC-001 through CC-099) were collected between August 2002 and February 2003 from 99 consenting Easter Island individuals by a practicing physician from Universidad de Valparaiso and were directly placed onto Whatman™ FTA™ cards (GE Healthcare, Waukesha, WI), which lyse cells as well as entangle, stabilize and preserve DNA. The FTA™ cards were stored at room temperature and DNA extraction of the samples was conducted in the summer of 2010.

Additional samples of 29 consenting Mapuche Native American individuals from the central region of Chile were obtained (RB-001 through RB-029) between August 2002 and February 2003 with IRB approval of Universidad de Chile. These samples were also collected by fingerstick and the blood was directly placed onto FTA™ cards.

All human specimen used in this study were covered under IRB protocol 2010-106.

Preliminary Study:

A multiplex of the nine Ramos *et al.* primer pairs was used to amplify three of the extracted Easter Island samples (CC-016, CC-028 and CC-056) along with a positive control using HL-60 cell line DNA and a negative control. The amplified products were then analyzed using the Agilent DNA 7500 kit on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA)(Appendix A). The Agilent 2100 Bioanalyzer uses microfluidics technology to separate

PCR-amplified products via electrophoresis and determines the length and quantity of each amplicon in a sample. An electropherogram or graph of the separated fragments in a sample is generated to view the fragments amplified in a PCR. The PCR-products of the three Easter Island samples contained very little of the targeted amplicons and more non-specific, shorter products compared to the positive control sample (Figure 2).

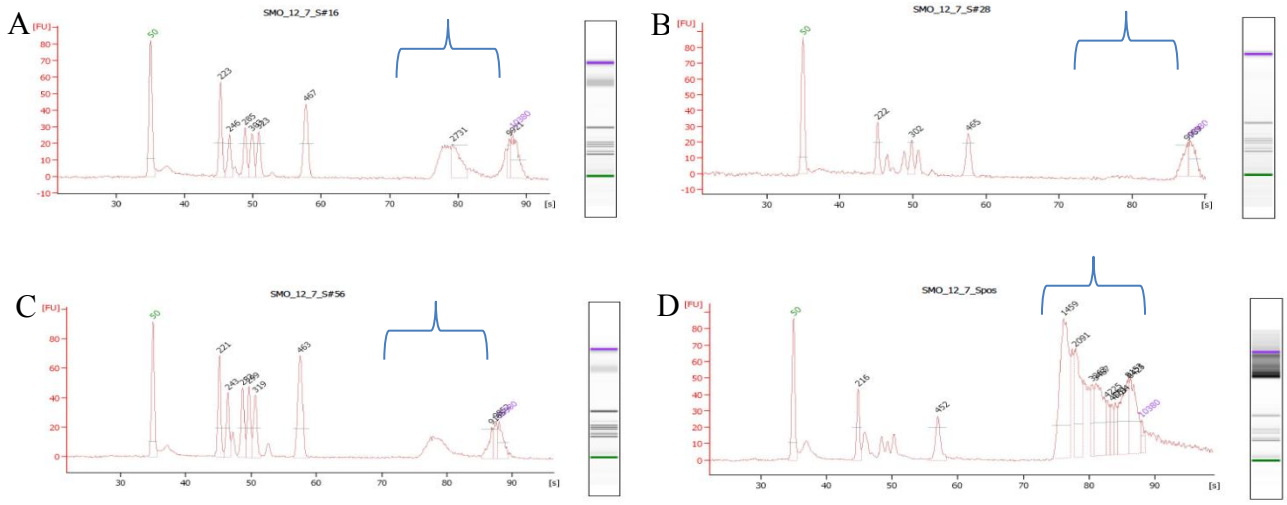


Figure 2: Electropherograms generated by the Agilent 2100 Bioanalyzer for the PCR products of three Easter Island samples amplified with the Ramos *et al.* primers. Fragment lengths are shown above each peak (in base pairs). A) CC-016; B) CC-028; C) CC-056; D) Positive control (HL-60). Little to none of the targeted fragments (1,500-3,000 bp) were amplified in the Easter Island samples (A-C) compared to the HL-60 positive control (D)(See blue brackets).

Direct amplification of the same three samples from the original FTA™ cards resulted in only slightly improved target amplicon quantity. Because the samples were placed directly onto FTA™ cards when they were collected and the mitochondrial genome is robust, it is not likely that the samples are degraded. Another explanation is that the FTA™ card matrix, which entangles and stabilizes DNA, keeps large fragments from being extracted or amplified. Based on this explanation, it is hypothesized that using a set of primers with shorter target amplicons

will successfully amplify the whole mitochondrial genome of these samples. A multiplex of PCR primers that successfully amplifies the mitochondrial genome of these Easter Island samples could also be used for full mitochondrial amplification of other FTA™ card samples that may not allow amplification of 1,500 bp or larger fragments.

Multiplex Development:

A literature search was conducted to find overlapping primers that would amplify smaller fragments of the complete mitochondrial genome to be used in conjunction with the nine Ramos *et al.* primer pairs. A study by Rieder *et al.* was found in which 24 primer pairs (1F/1R through 24F/24R) were developed that generated overlapping mtDNA amplicons between 765 and 1,162 base pairs in length (13). The study used BLAST to check the primers for specificity to the mitochondrial genome and used MITOMAP, a database of known polymorphisms in the mitochondrial genome, to look for variants at 3'-end positions of the primers that could affect the binding of the primers to the template DNA (13). Once the primers were determined to work in theory, Rieder *et al.* then tested each primer pair in singleplex reactions to determine that each pair amplified as expected (13).

Because the primers were found to amplify the whole mitochondrial genome in 24 separate amplification reactions by Rieder *et al.*, these same primers were obtained for this study to develop a set of multiplex reactions to generate a pool of amplicons that could then be used to develop a library for downstream next generation sequencing.

The primers were first tested in single primer pair amplification reactions to ensure they were all amplifying a single fragment of DNA at expected size in three samples: Easter Island sample CC-012 (on FTA™) and two positive control samples, TF1 (organically extracted buccal swab

DNA), and HL-60 (commercially available cell line DNA). PCR was conducted in 25 µL volume reactions consisting of 14.5 µL molecular grade water, 2.5 µL 10X PCR Buffer II, 2.5 µL Bovine Serum Albumin (1.6 µg/µL), 2.0 µL $MgCl_2$ (25mM), 1.0 µL dNTP mix (10mM), 0.5 µL of each primer (10 µM), 0.5 µL AmpliTaq Gold® (5 U/µL)(Life Technologies™, Carlsbad, CA), and 1 µL genomic DNA (1ng/µL). Each reaction was performed using the same set of thermal cycling conditions: 95°C for 11 minutes, 36 cycles of 95°C for 10 seconds, 61°C for 30 seconds, and 72°C for 30 seconds, and finally, 70°C for 10 minutes on an Eppendorf MasterCycler® pro S (Eppendorf International, Hamburg, Germany). Amplified products were evaluated on an Agilent 2100 Bioanalyzer using the Agilent DNA 1000 kit. The singleplex reactions generated single amplicons of the expected size at all primer pairs except for 18F/18R and 21F/21R that did not produce results. Primer pair 24F/24R amplified in the positive controls but not in the Easter Island sample. The original Ramos *et al.* forward primer 16488for was substituted for the 24F primer and amplification of the 24 fragment in the Easter Island sample was successful. New primers to amplify the 18 and 21 fragments were found in the National Institute of Science and Technology (NIST) Human Mitochondrial DNA – Amplification and Sequencing Standard Reference Materials (SRMs) 2392 and 2392-I (14). The primers, F11901/R12876 and F14189/R14924, were tested and amplified DNA fragments at their expected sizes.

Table 1: List of the primers that were used in this study to amplify the entire human mitochondrial genome. The primers were developed by Rieder *et al.* unless otherwise indicated (* Primers from NIST SRMs 2392 and 2392-I; ** Primers from Ramos *et al.*).

Primer Name	Sequence (5' - 3')	Position on mtDNA (bp)	Amplicon Size (bp)
1F	ctcctcaaagcaatacactg	592	
1R	tgctaaatccaccttcgacc	1430	838
2F	cgatcaacctcaccacctct	1226	
2R	tggacaaccagctatcacca	2026	800
3F	ggactaacccctataccttctgc	1830	

3R	ggcaggtcaatttcactggt	2688	858
4F	aaatcttaccccgcctgttt	2480	
4R	aggaatgccattgcgattag	3365	885
5F	tacttcacaaagcgccttcc	3150	
5R	atgaagaatagggcgaaggg	3980	830
6F	tggctcctttaacctctcca	3777	
6R	aaggattatggatgcggttg	4679	902
7F	actaattaatcccctggccc	4466	
7R	cctggggtgggttttgtatg	5443	977
8F	ctaaccggctttttgccc	5238	
8R	acctagaaggttgcctggct	6050	812
9F	gaggcctaacccctgtcttt	5835	
9R	attccgaagcctggtaggat	6661	826
10F	ctcttcgtctgatccgtcct	6450	
10R	agcgaaggcttctcaaatca	7334	884
11F	acgccaaaatccatttcact	7129	
11R	cgggaattgcatctgttttt	8114	985
12F	acgagtacaccgactacggc	7908	
12R	tgggtggttggtgtaaatga	8816	908
13F	tttccccctctattgatccc	8602	
13R	gtggccttggtatgtgcttt	9416	814
14F	cccaccaatcacatgcctat	9211	
14R	tgtagccgttgagttgtggt	10149	938
15F	tctccatctattgatgagggtct	9967	
15R	aattaggctgtgggtggttg	10858	891
16F	gccatactagtctttgccgc	10653	
16R	ttgagaatgagtgtgaggcg	11511	858
17F	tcactctcactgcccaagaa	11295	
17R	ggagaatgggggataggtgt	12095	800
F11901*	tgctagtaaccacgttctcctg	11901	
R12876*	gatatcgccgatacggttg	12876	975
19F	aaacaacccagctctccctaa	12551	
19R	tcgatgatgtggtctttgga	13526	975
20F	acatctgtacccacgccttc	13319	
20R	agaggggtcagggttcattc	14287	968
F14189*	acaaacaatgttcaaccagtaac	14189	
R14924*	aggcgtctggtgagtagtgc	14924	735
22F	tgaaacttcggctcactcct	14837	
22R	agctttgggtgctaatggtg	15997	1160
23F	tcattggacaagtagcatcc	15792	
23R	gagtggttaatagggtgatag	31	808
16488for**	ctgtatccgacatctggttcct	16488	
24R	aggctaagcgttttgagctg	794	875

Once primers that all amplified the target regions were confirmed, multiplex reactions were created to allow for full mitochondrial genome amplification in two PCR reactions. Primer Mix #1 consisted of the odd-numbered Rieder *et al.* primers and F14198/R14924. Primer Mix #2 contained the even-numbered Rieder *et al.* primers, F11901/R12876 and 16488for in place of 24F (Table 2, Figure 3). The primers were staggered so that those with overlapping target amplicons would be in separate reactions. By doing this, the smaller non-targeted overlap fragments observed in the initial multiplex would not be amplified.

Table 2: List of the primer pairs incorporated into each of the multiplex reactions.

Primer Mix #1	**Primer Mix #2**
1F/1R	2F/2R
3F/3R	4F/4R
5F/5R	6F/6R
7F/7R	8F/8R
9F/9R	10F/10R
11F/11R	12F/12R
13F/13R	14F/14R
15F/15R	16F/16R
17F/17R	F11901/R12876
19F/19R	20F/20R
F14189/R14924	22F/22R
23F/23R	16488for/24R

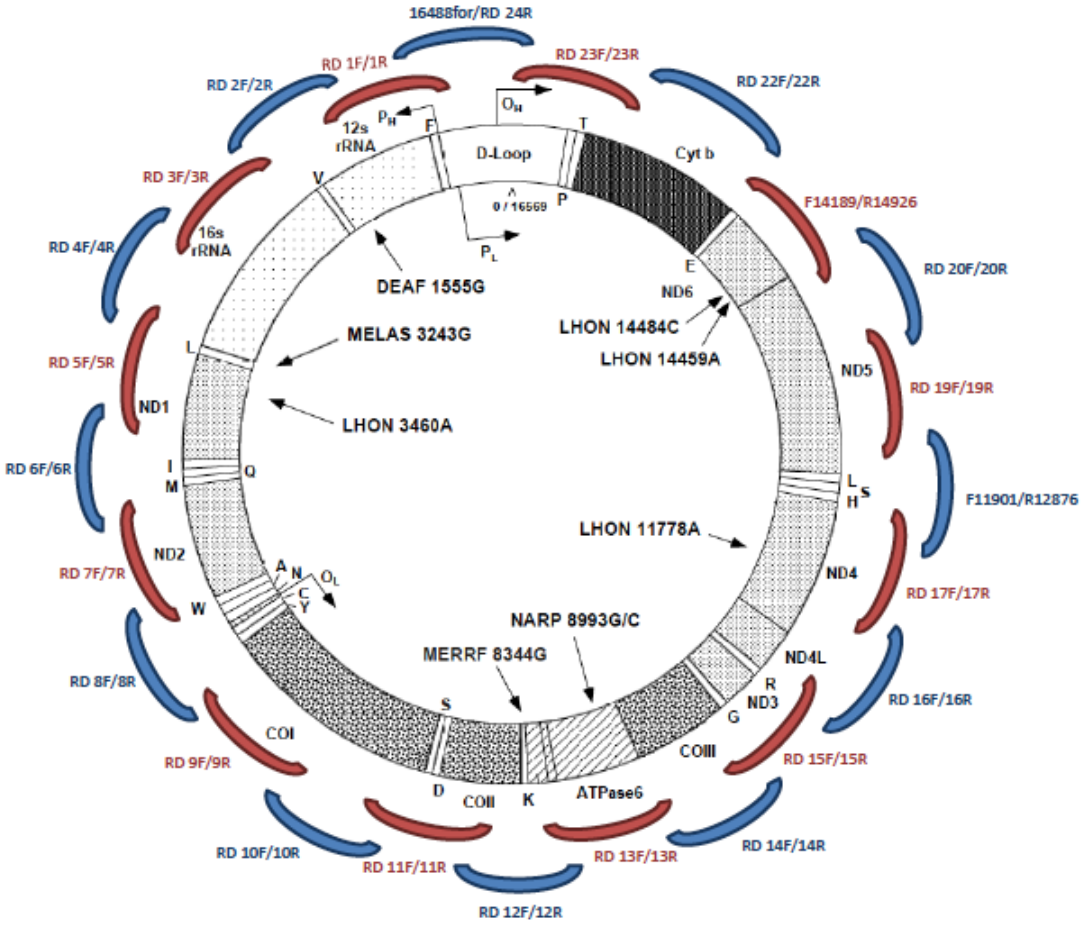


Figure 3: Depiction of the mitochondrial genome and the approximate locations of the targeted amplicons of each primer pair. Two multiplex PCR primer mixes were developed, one containing the odd primer pairs (red) and the other with the even primer pairs (blue). Adapted from http://www.transgenomic.co.uk/pd/MitoScreen.asp

Multiplex reactions used the Platinum® Multiplex PCR Master Mix (Applied Biosystems®, Foster City, CA). The three samples used previously (CC-097, TF1, and HL-60) were also used to test the multiplex reactions. Each reaction consisted of 25 µL 2X Platinum® Multiplex PCR Master Mix, 10 µL of either Primer Mix #1 and Primer Mix #2 (0.6 µM of each primer), 6 µL GC Enhancer, 4 µL molecular grade water, and 5 µL genomic DNA (1 ng/µL) for a total reaction volume of 50 µL. Thermal cycling conditions outlined in the Platinum® Multiplex PCR Master

Mix User Guide were followed: 95°C for 2 minutes, 35 cycles of 95°C for 30 seconds, 60°C for 90 seconds and 72°C for 60 seconds, and final extension at 72°C for 10 minutes on an Eppendorf MasterCycler® pro S.

An amplicon ladder was developed for each of the two multiplex reactions to compare the expected amplicon pools to the multiplex results. The odd amplicons from the HL-60 singleplex reactions were combined to form Amplicon Ladder #1. The even amplicons from the HL-60 singleplex reactions were combined to form Amplicon Ladder #2. The two ladders underwent post-PCR cleanup using the QIAquick® PCR Purification Kit (Qiagen® Inc., Valencia, CA)(Appendix B) and were then electrophoresed using the Agilent DNA 1000 kit on the Agilent 2100 Bioanalyzer with the multiplex PCR products to determine if the multiplex reactions amplified as expected.

Direct amplification of four Easter Island samples on FTA cards with the two multiplex reactions was performed to confirm the testing procedure. A Harris Uni-Core™ (Ted Pella, Inc., Redding, CA) device was used to take 1.2 mm punches from the FTA™ cards of CC-018, CC-023, CC-056 and CC-097. Two punches from each sample were taken, one for each multiplex reaction. The punches were processed with a series of washes in FTA™ Purification Reagent and TE buffer and then dried (Appendix C). Multiplex reaction mix, consisting of 25 μL 2X Platinum® Multiplex PCR Master Mix, 10 μL of either the Primer Mix #1 or Primer Mix #2, 6 μL GC Enhancer and 9 μL of molecular grade water, was added to each FTA™ sample. The same Platinum® Multiplex thermal cycling parameters for were used for these amplification reactions (Appendix D). The PCR products were electrophoresed on the Agilent 2100 Bioanalyzer along with the amplicon ladders to determine whether amplification was successful.

Three Mapuche Native American samples were selected to undergo direct amplification with the newly developed multiplex reactions as well. These samples were RB-004, RB-009 and RB-021. The same procedure as that used for direct amplification of the Easter Island samples was used for these samples. The PCR products were electrophoresed on the Agilent 2100 Bioanalyzer along with amplicon ladders #1 and #2 to observe whether amplification of the targeted fragments was successful.

Library Preparation and Sequencing:

Once amplicon pools for the four Easter Island and three Mapuche samples were obtained, preparation of libraries for next generation sequencing on the Ion Torrent™ Personal Genome Machine (PGM™) Sequencer® (Life Technologies™, Carlsbad, CA) was conducted. In addition to the seven pilot samples, the amplicon pool of HL-60 underwent library preparation to act as a positive control. The PGM™ is an ion semiconductor instrument that acts as a pH meter and can detect when a hydrogen atom is released as a nucleotide is being added to the elongating strand of DNA in a sequence reaction (15). If the specific base added to the reaction is incorporated into the elongating strand, the pH of the solution will change and that base will be recorded as the next base in the sequence (15).

Preparation of the amplicon libraries was completed using the NEBNext® Fast DNA Fragmentation & Library Prep Set for Ion Torrent™ (New England Biolabs®, Inc., Ipswich, MA). Because the amplicons were approximately 700-1100 base pairs in length, they were enzymatically fragmented so that 200 base pair fragments could be selected to be run on the PGM™. Each fragmentation reaction consisted of 500 ng of amplicons, 1 μL of $MgCl_2$, 2 μL NEBNext® DNA Fragmentation Reaction Buffer, 1.5 μL NEBNext® DNA Fragmentation Master

Mix, and a variable volume of molecular grade water to bring the total reaction volume to 20 µL. Tests were performed to determine the optimal thermal cycler parameters for the reaction. The fragmentation protocol in the NEBNext® Fast DNA Fragmentation & Library Prep Set kit called for 20 minutes at 25°C (enzymatic activity) and 10 minutes at 70°C (enzymatic deactivation). Because this protocol was designed for fragmenting genomic DNA, new parameters were established to fragment the 700-1100 base pair amplicons in this study. The optimal parameters were determined to be 25°C for 8.5 minutes followed by 70°C for 10 minutes (Appendix E). End repair of the fragmented amplicons was performed simultaneously by enzymes in the reaction mix.

After fragmentation, NEXTflex™ DNA barcode adapters (Bio Scientific Corporation, Austin, TX) were ligated to the amplicons. These adapters are about 50 base pairs in length. The barcodes on the adapters allow for multiple samples to be sequenced in a single PGM™ run. The ligation reaction took place using reagents from the NEBNext® Fast DNA Fragmentation & Library Prep Set for Ion Torrent™ kit. The following reagents were added to the fragmentation reaction for a total reaction volume of 60 µL: 3 µL molecular grade water, 10 µL T4 DNA Ligase Buffer for Ion Torrent, 1 µL *Bst* DNA Polymerase, 10 µL NEXTflex™ DNA Barcode Adapter (1-8, one for each sample), 10 µL NEXTflex™ DNA P1 Adapter and 6 µL T4 DNA Ligase. The ligation reaction was conducted on a thermal cycler at 25°C for 15 minutes followed by 65°C for 5 minutes (Appendix F). The ligated DNA fragments were purified using the QIAquick® PCR Purification Kit (Appendix B).

The adapter ligated samples were then run on the Lonza FlashGel™ System (Lonza Rockland, Inc., Rockland, ME)(Appendix G) to select base pair fragments approximately 250 bp in length, the optimal size read for the PGM™. The size-selected, barcode adapter ligated DNA

fragments then underwent PCR amplification. Each reaction consisted of the size-selected adapter ligated DNA (approximately 30-40 μL), 10 μL of NEXTflex™ Primer Mix, 50 μLNEBNext High-Fidelity 2X PCR Master Mix and a variable volume of molecular grade water to bring the total reaction volume to 100 μL. The PCR thermal cycling conditions included 98°C for 30 seconds, 5 cycles of 98°C for 10 seconds, 58°C for 30 seconds and 72°C for 30 seconds, and finally, 72°C for 5 minutes (Appendix H). The PCR products were purified using Agencourt® AMPure® XP Reagent (Beckman Coulter, Inc., Indianapolis, IN) (Appendix I).

The quantity of the library was assessed using quantitative PCR (qPCR) with the Ion Library Quantitation Kit (Life Technologies™, Carlsbad, CA) (16). Dilutions of the libraries were made based on the qPCR results and the libraries for each sample were pooled together. The pooled library then underwent emulsion PCR on the Ion OneTouch™ System (Life Technologies™, Carlsbad, CA) and the combine library pool for all samples was injected onto the PGM™ chip for sequencing (17-18).

Data Analysis:

The resulting sequence reads for each sample were aligned to the mitochondrial revised Cambridge Reference Sequence (rCRS)(GenBank accession number NC_012920) and assessed to determine the consensus sequence, the level of coverage of the mitochondrial genome, identify polymorphic sites, and evaluate heteroplasmy detected. FASTQ data files containing the sequence reads from the PGM™ were converted to FASTA format using an online bioinformatics analysis platform called the Galaxy Project (galaxyproject.org). The FASTA files were then uploaded into NextGENe® (SoftGenetics®, State College, PA) for sequence analysis (Appendix J). The curated consensus sequences were uploaded into MEGA5 (Molecular

Evolutionary Genetics Analysis software) (19) to conduct an evaluation of the phylogenetic placement of the test samples with regard to a set of reference sequences. Thirty previously-published full mitochondrial Oceanian (20-24) and six previously-published sequences of native Argentinians (25) were uploaded into the MEGA project to compare to the Easter Island and Mapuche sequences (Appendix K). The Oceanian samples included individuals from Papua New Guinea, Samoa, the Cook Islands, Tonga, the Philippines, Taiwan and East Malaysia. MEGA was used to align all of the sequences using MUSCLE (Multiple Sequence Comparison by Log-Expectation). Genetic distances between individuals were calculated under the Tamura-Nei model (26). A neighbor joining tree with bootstrapping (500 iterations) depicting the relative genetic distances between the samples was created to generate an estimate of the genetic placement of the pilot samples.

CHAPTER III

RESULTS

Singleplex Reactions:

Electropherograms of the 24 singleplex reactions for CC-012, TF1, and HL-60 showed that each primer pair amplified its respective targeted size amplicon. Figure 4 shows the singleplex reaction electropherograms of HL-60. These amplicons were combined to form Amplicon Ladders #1 and #2 that were electrophoresed with the multiplex reactions. The singleplex electropherograms for CC-012 and TF1 are provided in Appendix L.

Multiplex Development:

The amplicon products of the first multiplex reactions for samples CC-097, TF1 and HL-60 matched that of the expected amplicon products based on constructed Amplicon Ladders #1 and #2 (Figures 5 and 6, remaining electropherograms located in Appendix M). As many of the amplicons are similar in size (see Table 1), the Agilent 2100 Bioanalyzer could not completely resolve each amplicon. Therefore, some peaks on the electropherograms include multiple amplicons. The ladders were much less concentrated than the samples (as evidenced by the internal sizing/concentration standards, labeled in green and purple in the electropherograms), however, the peaks present in the ladders were also present in the samples at the same relative concentration. This indicates that all of the primer pairs amplified their respective fragments.

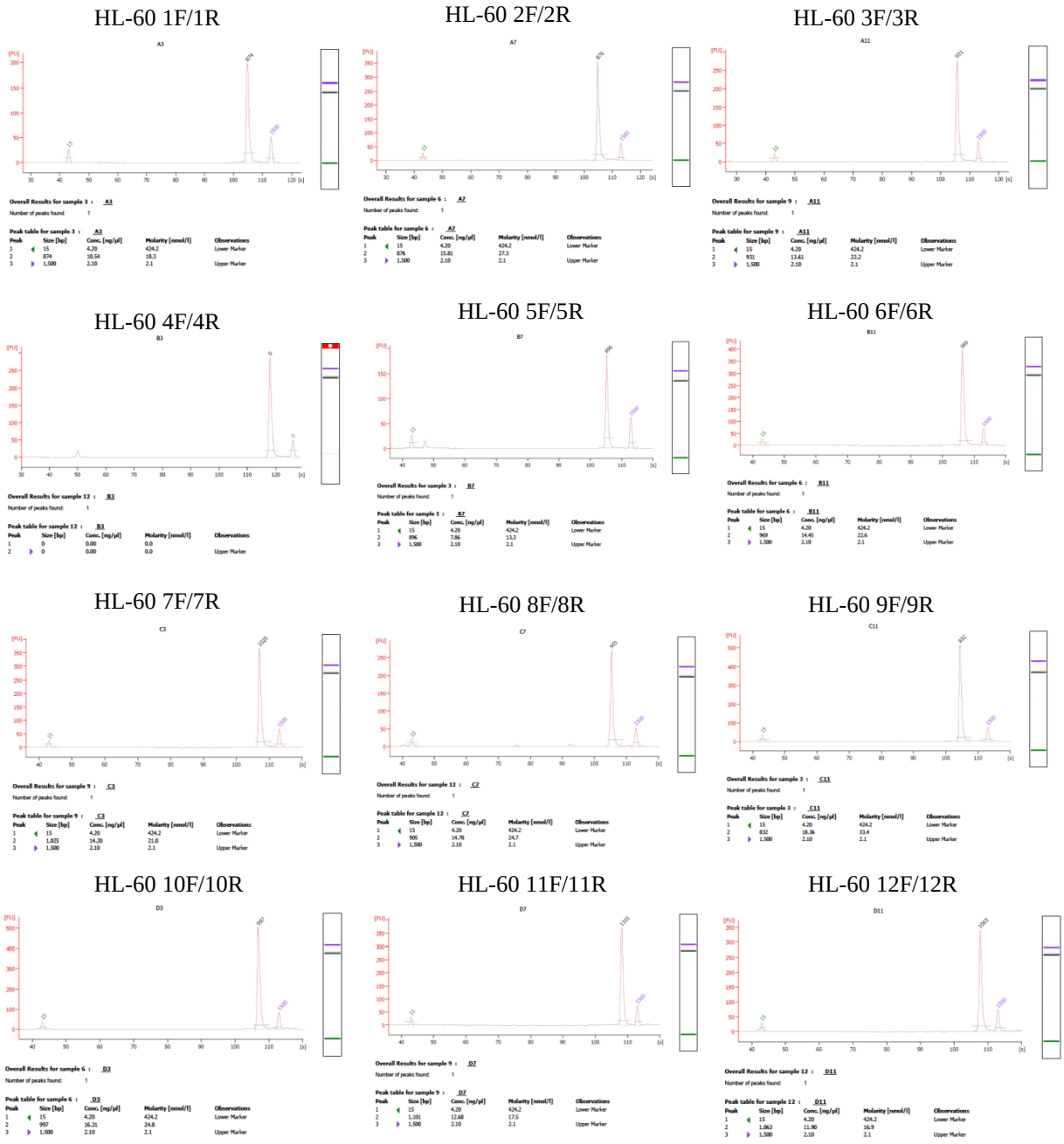
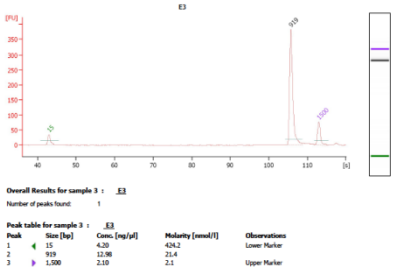
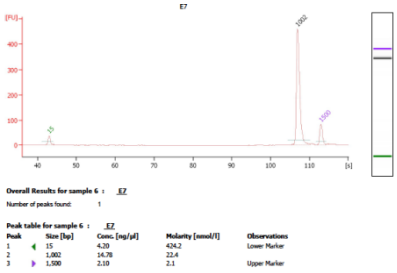


Figure 4: Electropherograms of the singleplex reactions for positive control samples HL-60. The electropherogram for primer pair 4F/4R shows that the targeted amplicon was produced, but electrophoretic issues kept the amplicon from being sized or quantified. (Electropherograms of the singleplex reactions for CC-012 and TF1 are provided in Appendix L.)

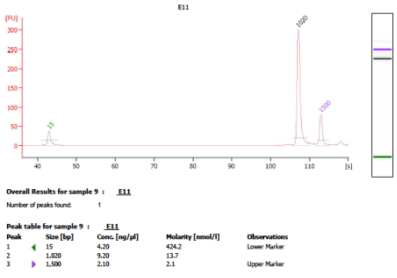
HL-60 13F/13R



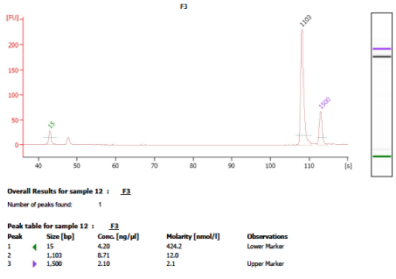
HL-60 14F/14R



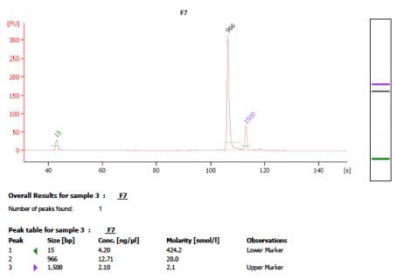
HL-60 15F/15R



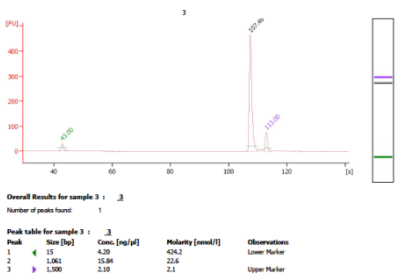
HL-60 16F/16R



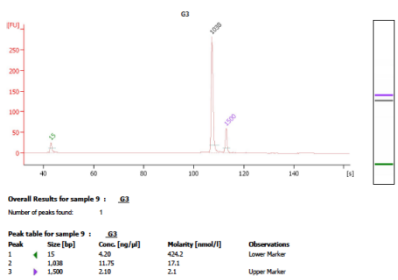
HL-60 17F/17R



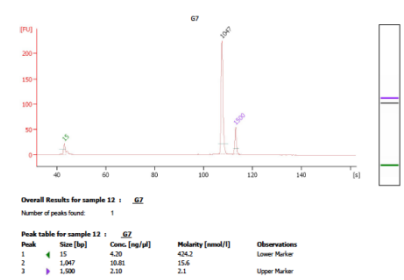
HL-60 F11901/R12876



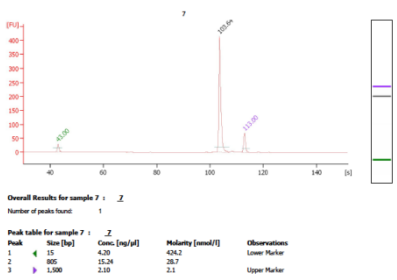
HL-60 19F/19R



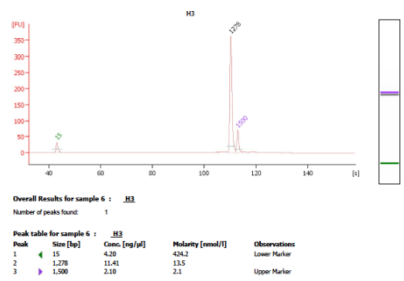
HL-60 20F/20R



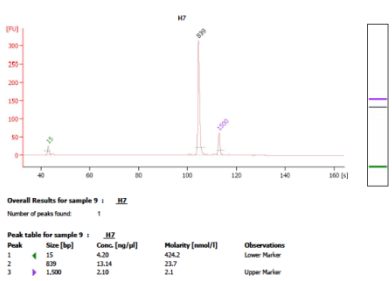
HL-60 F14189/R14924



HL-60 22F/22R



HL-60 23F/23R



HL-60 16488for/24R

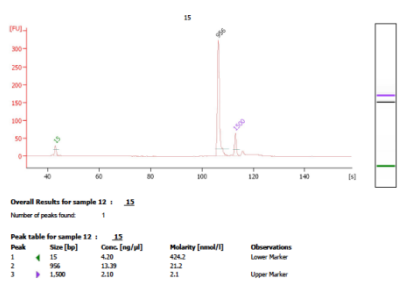
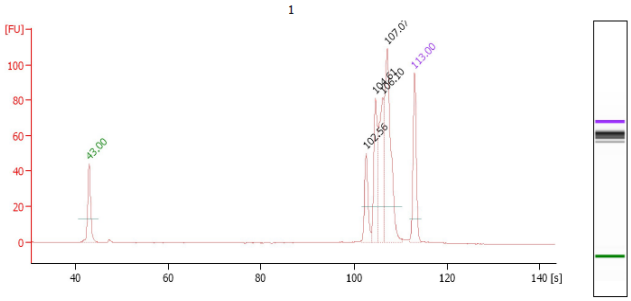


Figure 4 continued: Electropherograms of the singleplex reactions for positive control samples HL-60. (Electropherograms of the singleplex reactions for CC-012 and TF1 are provided in Appendix L.)

Amplicon Ladder #1



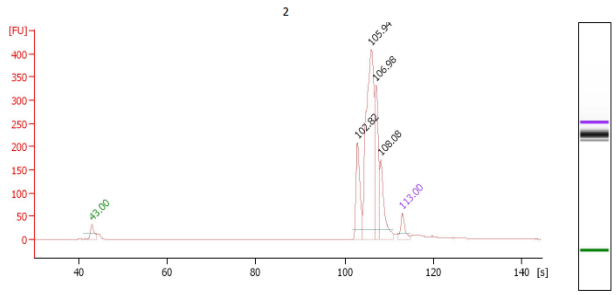
Overall Results for sample 1 : 1

Number of peaks found: 4

Peak table for sample 1 : 1

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	751	1.35	2.7	
3	853	2.33	4.1	
4	956	2.65	4.2	
5	1,027	4.85	7.2	
6	1,500	2.10	2.1	Upper Marker

CC-097 Primer Mix #1- Extracted



Overall Results for sample 2 : 2

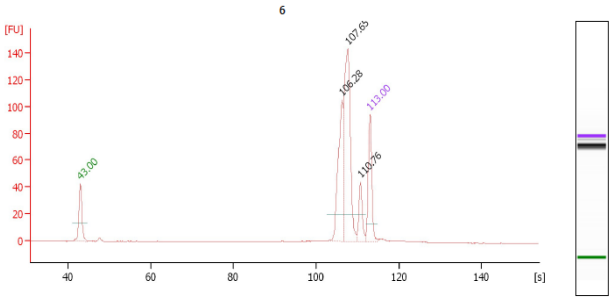
Number of peaks found: 4

Peak table for sample 2 : 2

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	764	8.35	16.6	
3	945	28.82	46.2	
4	1,021	11.18	16.6	
5	1,108	7.04	9.6	
6	1,500	2.10	2.1	Upper Marker

Figure 5: Electropherograms of Amplicon Ladder #1 and extracted sample CC-097 amplified with Primer Mix #1. All of the amplicons present in the ladder were also present in the amplified sample. (Electropherograms of the other extracted samples amplified with Primer Mix #1 are provided in Appendix M.)

Amplicon Ladder #2



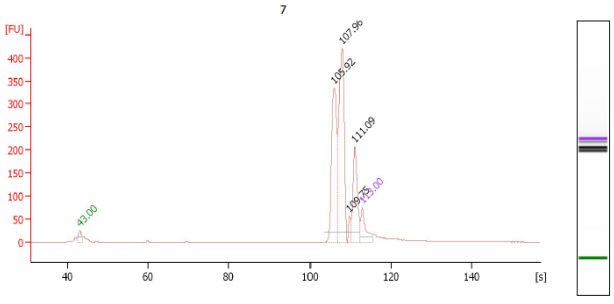
Overall Results for sample 6 : 6

Number of peaks found: 3

Peak table for sample 6 : 6

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	969	4.02	6.3	
3	1,074	6.47	9.1	
4	1,322	1.12	1.3	
5	1,500	2.10	2.1	Upper Marker

CC-097 Primer Mix #2 - Extracted



Overall Results for sample 7 : 7

Number of peaks found: 4

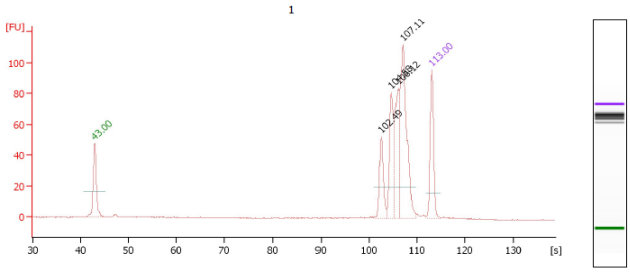
Peak table for sample 7 : 7

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	944	11.64	18.7	
3	1,098	13.77	19.0	
4	1,241	0.82	1.0	
5	1,347	5.44	6.1	
6	1,500	2.10	2.1	Upper Marker

Figure 6: Electropherograms of Amplicon Ladder #2 and extracted sample CC-097 amplified with Primer Mix #2. Amplicons present in the ladder were also present in the amplified sample. (Electropherograms of the other extracted samples amplified with Primer Mix #2 are provided in Appendix M.)

Since the extracted samples amplified as expected with the multiplex primer mixes, amplification directly from the FTA™ cards was conducted to determine whether similar results could be obtained with this method. Amplicon pools generated directly from the FTA™ cards of samples CC-018, CC-023, CC-056, CC-097, RB-004, RB-009 and RB-021 also matched the expected products observed in the amplicon ladders (Figures 7 and 8). The successful direct amplification of the samples with the two primer mixes results in a more time-efficient method for creating the amplicon pools than having to perform extractions on the FTA™ cards prior to amplifying.

Amplicon Ladder #1



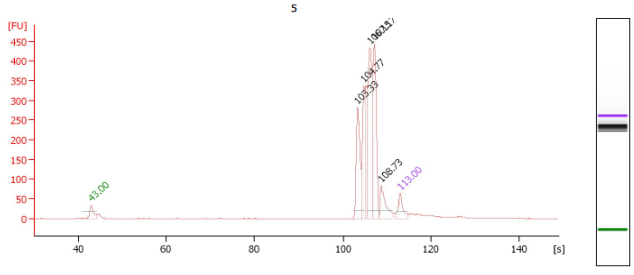
Overall Results for sample 1 : 1

Number of peaks found: 4

Peak table for sample 1 : 1

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	752	1.38	2.8	
3	856	2.32	4.1	
4	964	2.71	4.3	
5	1,037	5.01	7.3	
6	1,500	2.10	2.1	Upper Marker

CC-097 Primer Mix #1 - Direct Amplification



Overall Results for sample 5 : 5

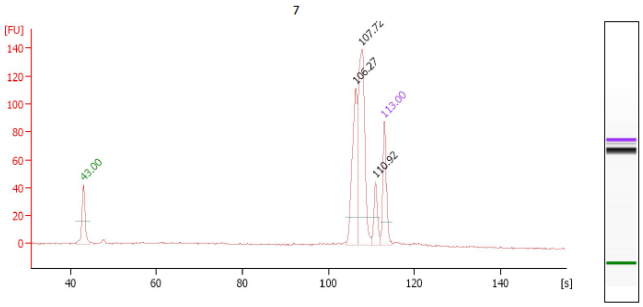
Number of peaks found: 5

Peak table for sample 5 : 5

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	793	8.42	16.1	
3	869	7.33	12.8	
4	966	14.60	22.9	
5	1,042	11.86	17.2	
6	1,165	3.12	4.1	
7	1,500	2.10	2.1	Upper Marker

Figure 7: Electropherograms of Amplicon Ladder #1 and sample CC-097 directly amplified with Primer Mix #1. Amplicons present in the ladder were also present in the amplified sample. (Remaining electropherograms of the other Easter Island and Mapuche samples amplified directly from FTA™ punches with Primer Mix #1 are provided in Appendix N.)

Amplicon Ladder #2



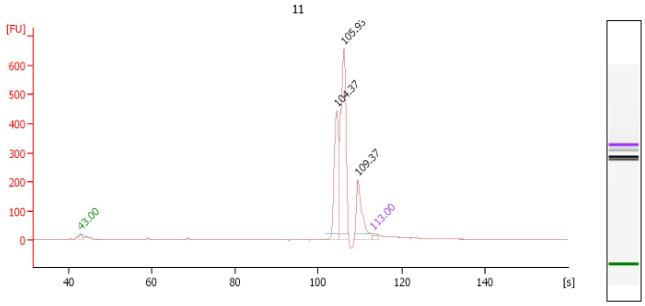
Overall Results for sample 7 : 7

Number of peaks found: 3

Peak table for sample 7 : 7

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	974	4.49	7.0	
3	1,085	6.73	9.4	
4	1,337	1.19	1.3	
5	1,500	2.10	2.1	Upper Marker

CC-097 Primer Mix #2 - Direct Amplification



Overall Results for sample 11 : 11

Number of peaks found: 3

Peak table for sample 11 : 11

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	844	42.45	76.2	
3	950	66.30	105.7	
4	1,215	23.76	29.6	
5	1,500	2.10	2.1	Upper Marker

Figure 8: Electropherograms of Amplicon Ladder #2 and sample CC-097 directly amplified with Primer Mix #2. Amplicons present in the ladder were also present in the amplified sample. (Remaining electropherograms of the other Easter Island and Mapuche samples amplified directly from FTA™ punches with Primer Mix #2 are provided in Appendix N.)

Library Preparation and Sequencing:

The pooled amplicons from the two multiplexes for each sample were fragmented enzymatically for 8.5 minutes at 25°C and the enzyme was deactivated for 10 minutes at 70°C. The resulting electropherogram showed a smear of fragments ranging in size from about 15 to 1500 base pairs (Figure 9). The target size fragments for sequencing were approximately 200 – 300 base pairs. This size fragment was selected for using the LonzaFlashGel™ System. The rest of the fragments not meeting the size requirement were not used in the downstream library preparation.

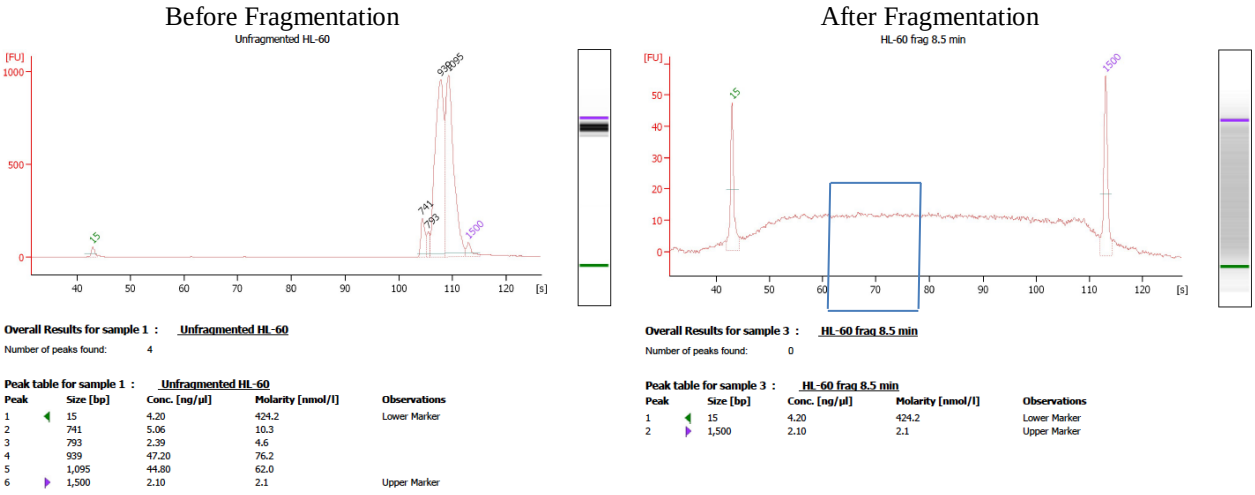


Figure 9: Electropherograms of the pooled amplicons of positive control HL-60 before and after enzymatic fragmentation using the NEBNext® Fast DNA Fragmentation & Library Prep Set for Ion Torrent™ kit. The blue box indicates the fragment size range that was selected for in the size selection step with the Lonza FlashGel™ System (~200-300 bp).

After the sequencing was performed, the resulting data was uploaded into NextGENe® software for alignment to the rCRS and to determine the consensus sequence, coverage level and heteroplasmy levels for each sample. Table 3 displays the total number of reads, the average size of the sequenced reads and the coverage level. Combined, the samples had over 150,000 reads. The average read length for each sample was smaller than the targeted length of approximately 250 base pairs, but overlap inherent in the design of the enrichment amplification of the mitochondrial genome permitted high coverage. Each sample had an average coverage of at least 1,100X, and at least 98% of the mitochondrial genome was covered by reads.

Table 3: Information on the sequencing results obtained for each sample from the NextGENe® software. All samples produced over 150,000 reads. The average read length was smaller than the targeted size (~250 bp), but all samples had over 1,100X average coverage and over 98% of the mitochondrial genome was covered in each sample.

Sample	Total Reads	Average Read Length (bp)	Average Coverage	No. Bases Covered	Percent Coverage
CC-018	234,098	103	1397	16305	98.406
CC-023	179,028	112	1159	16366	98.775
CC-056	334,387	120	2260	16550	99.885
CC-097	364,961	94	1889	16548	99.873
HL-60	262,478	117	1706	16569	100
RB-004	417,475	111	2569	16557	99.928
RB-009	423,078	121	2832	16549	99.879
RB-021	440,193	95	2290	16533	99.783

Complete coverage of the mitochondrial genome was only obtained in the positive control HL-60. Figure 10 displays the relative coverage observed across the mitochondrial genome of each sample and the relative positions of the targeted amplicons for each primer pair used in the multiplex reactions developed in this study. Coverage dropped out between the 7,500 and 8,000 base pair positions in samples CC-018 and CC-023. This same area in the other samples contained low coverage. This area corresponded to the position of the targeted amplicon of primers 11F and 11R. Another area of low or no coverage was observed between positions 13,500 and 14,250. Those positions corresponded to the area in which primers 20F and 20R are expected to amplify. Complete loss of coverage in this area was observed in all of the samples except for the positive control HL-60. Loss of coverage in these two areas could be due to inefficiency of primer binding due to a primer binding site mutation in the targeted genomes. Another possibility is that the concentrations of these primers in the primer mixes were not high enough to produce a significant number of amplicons to go into library preparation.

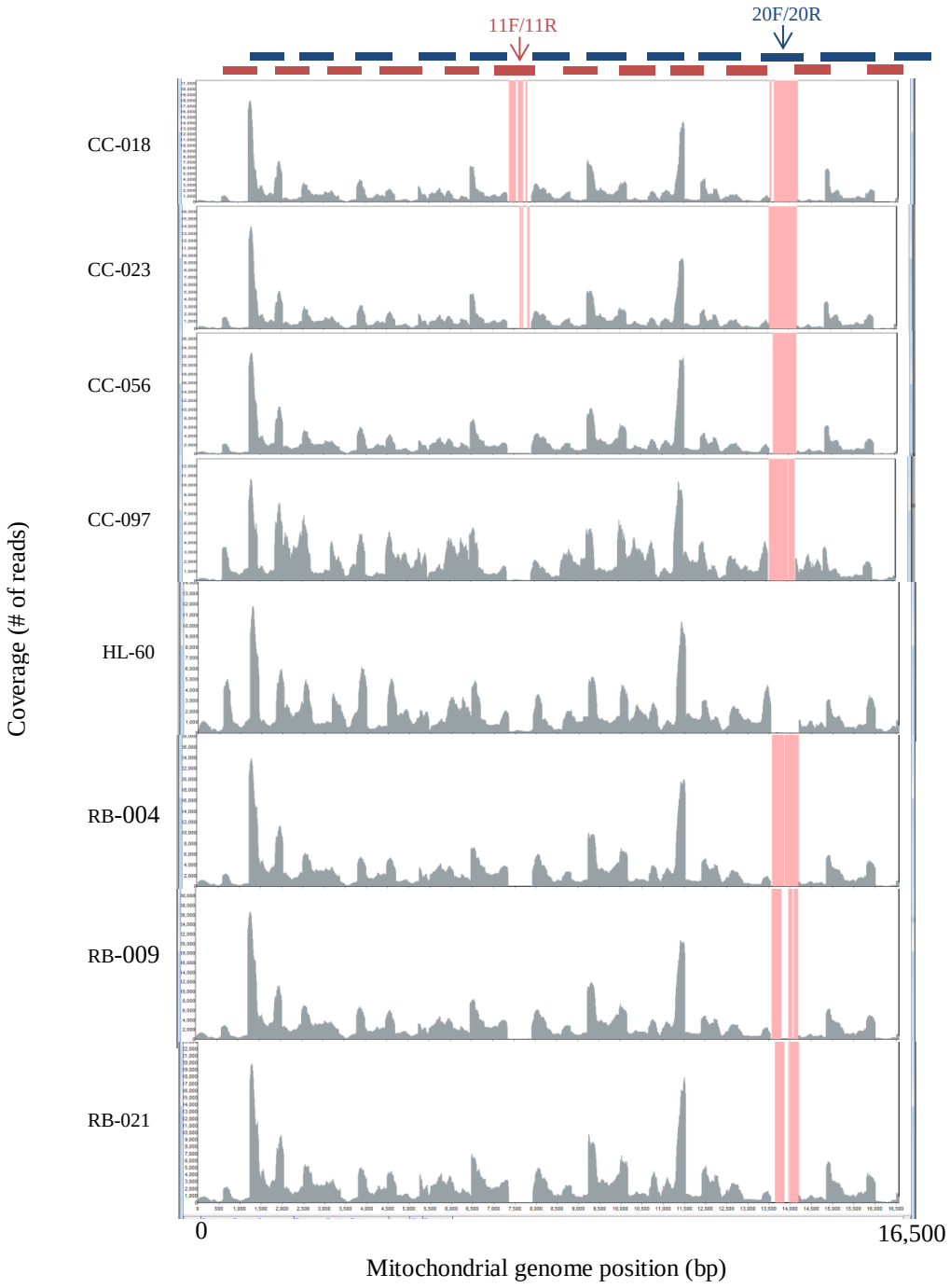


Figure 10: Coverage curves for each sample depicting the amount of coverage observed across each entire mitochondrial genome. The pink bars indicate areas where very low coverage or complete read dropout occurred. The approximate positions of the targeted amplicons for each primer pair used in the multiplex reactions are displayed at the top of the figure. Little to no coverage was observed in the areas expected to be amplified by primer pairs 11F/11R and 20F/20R.

The NextGENe® software was used align the sequencing reads and derive a consensus sequence for each sample. Positions in the sequence that contained read coverage with two nucleotide calls with the minor nucelotide at least 20 percent of the major nucleotide, were determined to be heteroplasmic sites. Heteroplasmy was detected in all of the samples but one, RB-004 (Table 4). The low level nucleotides were observed at levels ranging from 25.9% in sample CC-023 to 89.2% in the positive control HL-60.

Table 4: The number of heteroplasmic sites observed in the consensus sequences of the samples. Heteroplasmy was detected in all of the samples except RB-004. Nucleotide ratios (minor/major) ranged from 0.259 in sample CC-023 to 0.892 in positive control sample HL-60.

Sample	# Heteroplasmic sites	Nucleotide ratios (Minor/Major)
CC-018	10	0.281 – 0.828
CC-023	5	0.259 – 0.592
CC-056	4	0.391 – 0.477
CC-097	4	0.269 – 0.333
HL-60	1	0.892
RB-004	0	---
RB-009	6	0.262 – 0.467
RB-021	5	0.262 – 0.589

The consensus sequence of the positive control HL-60 matched that of the published HL-60 sequence in the NIST SRM 2393-I (14). The consensus sequences of the pilot samples were exported from the NextGENe® software in a FASTA file format and, uploaded into MEGA5 along with the previously-published full mitochondrial genome sequences from Oceanian populations and samples from Argentina. Forty-three sequences were aligned to the rCRS. Because full coverage of the sequenced samples was not obtained, missing data was treated as a complete deletion across all of the sequences. The C-stretch homopolymeric region located

around the 310 bp position of the mitochondrial genome was also removed from analysis from each sample because it is known to cause sequencing issues. A neighbor joining phenogram was created depicting the relative genetic distances between each individual (Figure 11). Bootstrap confidence estimates based on 500 iterations are depicted at the tree nodes.

The phenogram displayed two main clusters. The first cluster contained the Oceanians from Samoa, Cook Islands, Tonga, Taiwan and all four Easter Island samples. This cluster also contained some of the individuals from Papua New Guinea, the Philippines and Borneo. The second cluster included the Mapuche and Argentinians as well as the other individuals from Papua New Guinea, the Philippines and Borneo. These pilot results indicate that Easter Islanders are more closely affiliated with Oceanians than with mainland Chileans. It also suggests that gene flow between Oceanians and South American natives has occurred as has been previously reported (27).

To ensure that the areas that were ignored across all 43 sequences due to loss of sequencing coverage in the pilot samples did not contain phylogenetically informative sites that would affect phenogram topology, a neighbor joining tree containing complete sequences of the Oceanian and Argentinian samples was generated (Figure 12). If the Oceanian and Argentinian individuals clustered similarly in both neighbor joining trees, then the lost coverage areas would be determined to not affect the genetic distance results between the Easter Island, mainland Chilean and Oceanian samples depicted in Figure 11. The second phenogram provided similar clustering of the Oceanian and Argentinian individuals to that in Figure 11, indicating that the omitted regions did not contain any key phylogenetic sites or affect phenogram topology.

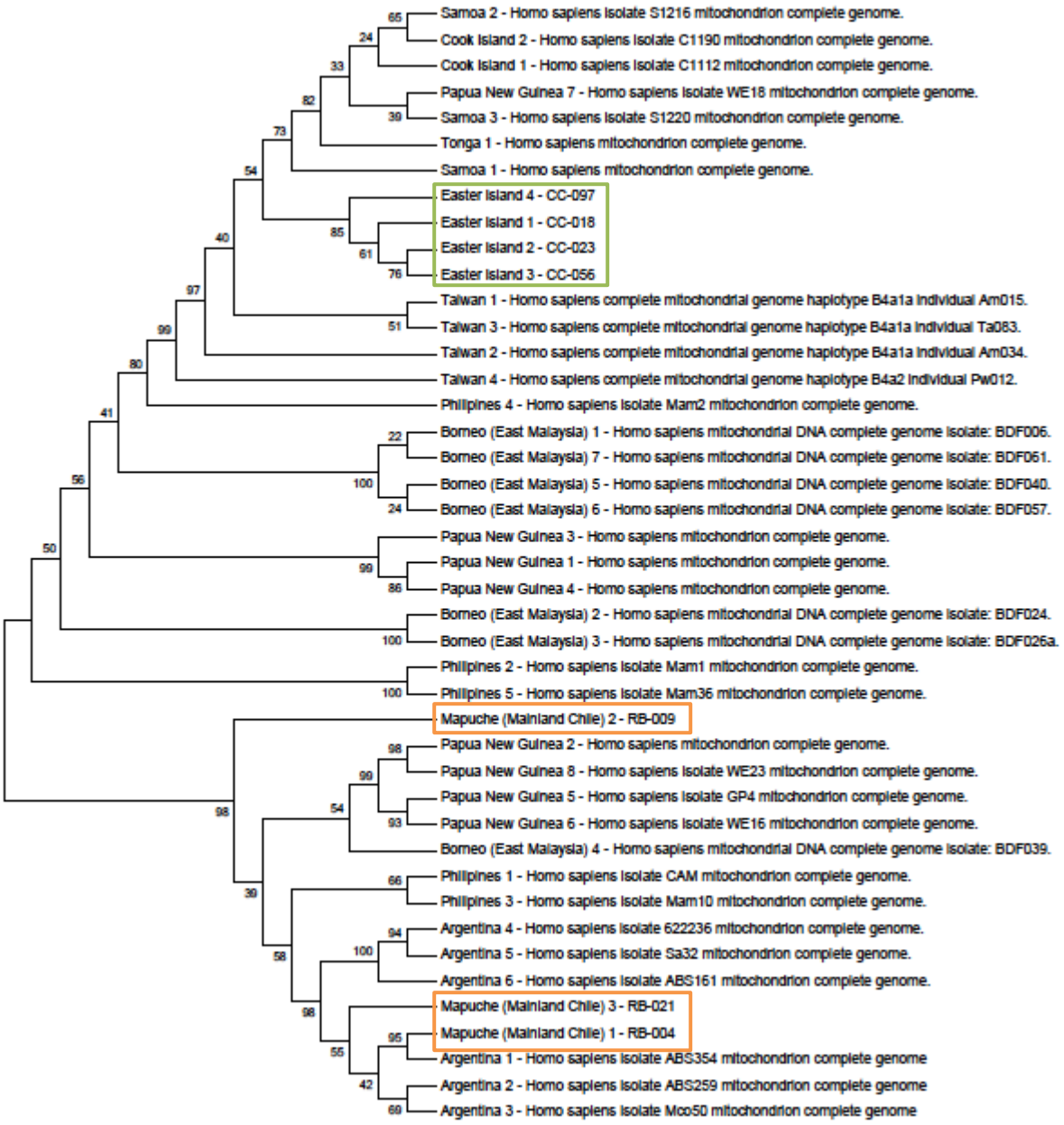


Figure 11: Neighbor joining tree phenogram of the Easter Island (green box) and Mapuche (orange boxes) samples sequenced in this study as well as the previously-published Oceanian and Argentinian full mitochondrial sequences. The phenogram shows two main clusters: one containing the Easter Islanders and the majority of the Oceanians and the other containing the South American natives and the rest of the Oceanians. These results indicate that Easter Islanders are more closely related to Oceanians, but that gene flow between the South American native and Oceanian populations has occurred.

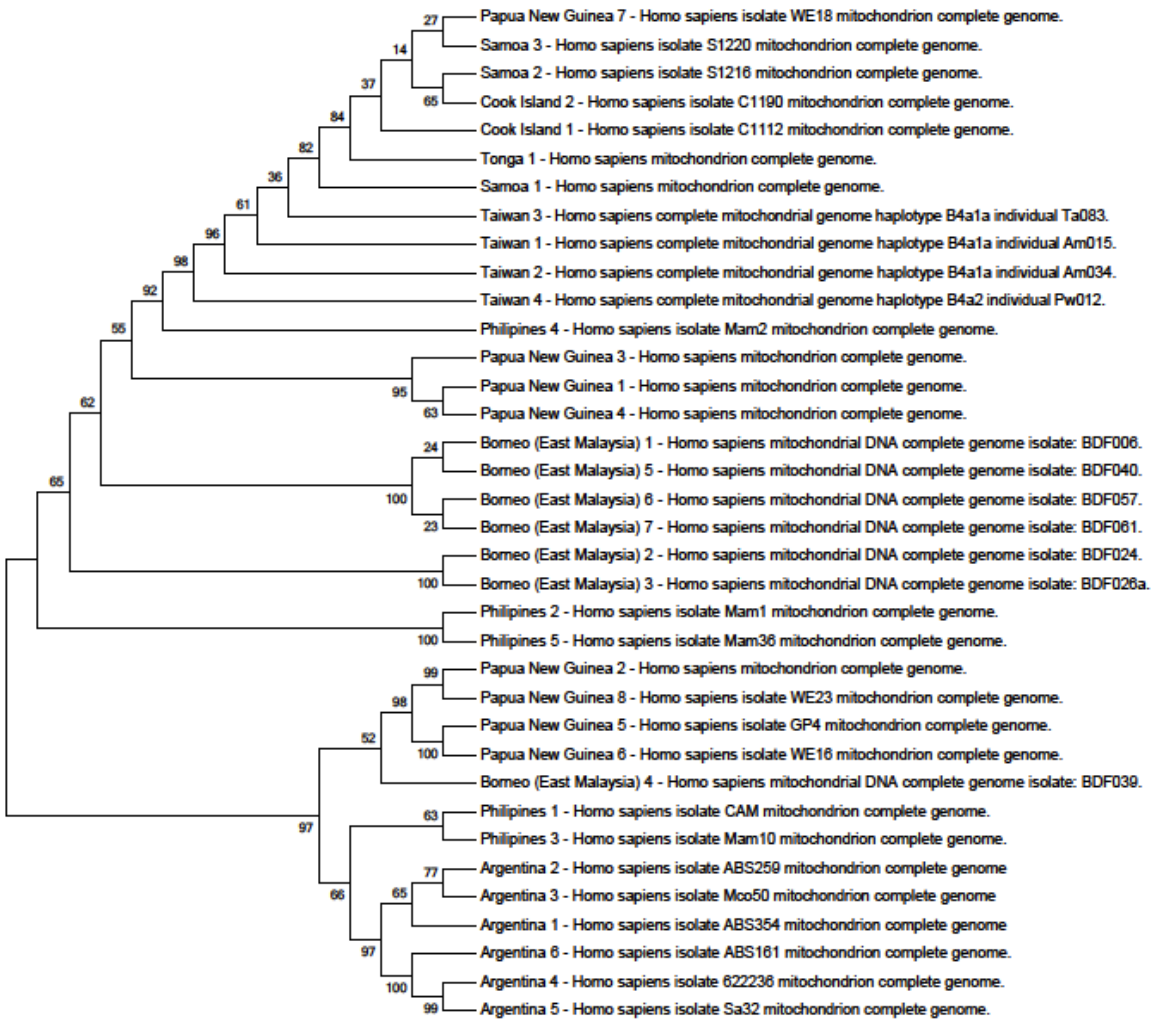


Figure 12: Neighbor joining tree phenogram of the previously-published Oceanian and Argentinian mitochondrial sequences including the areas that had been deleted from analysis in the first phenogram (Figure 11). Because these samples clustered similarly in both phenograms, it can be determined that the lost coverage in the samples sequenced for this study did not affect the genetic distance results of the comparison of Easter Island, Oceanian and mainland Chilean individuals depicted in Figure 11.

Finally, a maximum parsimony tree with bootstrapping (500 iterations) was constructed containing the 43 sequences with an individual from South Africa (GenBank accession number EF_184592) used as the root for the tree (Figure 13). The parsimony tree depicts evolutionary relationships between the Easter Island, Oceanian, mainland Chilean, and Argentinian

individuals. Similar clustering of individuals to that of the neighbor-joining tree (Figure 11) was observed in the maximum parsimony tree. Easter Islanders clustered with most of the Oceanians, while mainland Chileans clustered with Argentinians and the rest of the Oceanians.

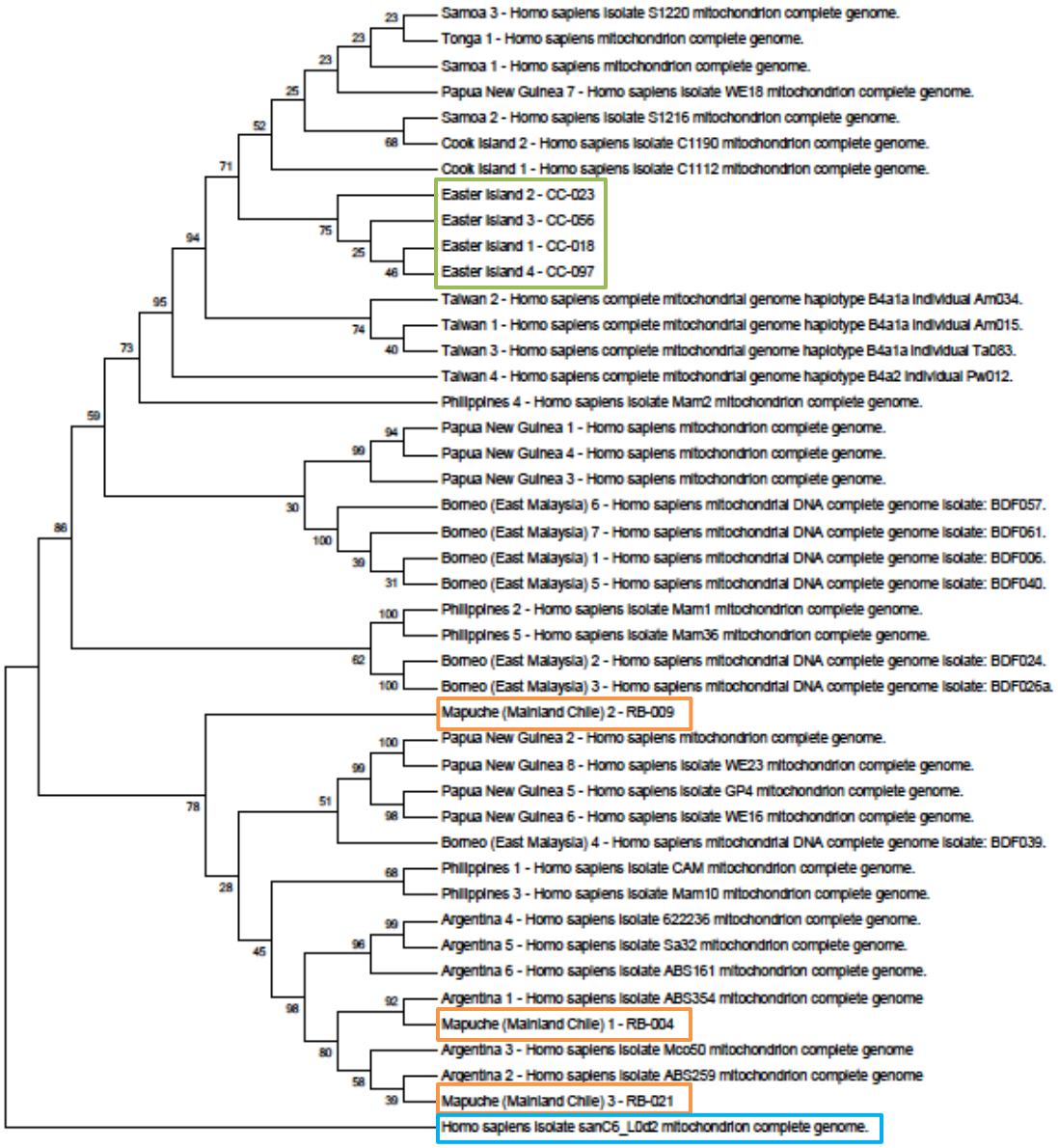


Figure 13: Maximum parsimony tree depicting the evolutionary relationships between the sequenced Easter Island (green box) and Mapuche (orange boxes) samples as well as the previously-published Oceanian and Argentinian mitochondrial sequences. An individual from South Africa was used as the root for the tree (blue box). Similar clustering was observed in the maximum parsimony tree and neighbor joining tree (Figure 11).

CHAPTER IV

DISCUSSION

The purpose of this study was to develop a multiplex that could amplify the entire mitochondrial genome so that the resulting amplicons could be used as a target enriched library for next generation sequencing. The multiplex reactions developed in this study amplified the samples extracted from FTA™ cards that did not amplify with the nine primer pair multiplex developed from the Ramos *et al.* primer set. Therefore, the hypothesis that targeting smaller amplicons of the mitochondrial genome than those targeted by the Ramos *et al.* primers would result in successful amplification was supported. These multiplexes were successful in producing an amplicon pool with direct amplification from the FTA™ card samples. Direct amplification of the samples using the newly developed multiplexes allows for a more efficient method for creating a target enriched amplicon pool.

The amplicons produced from the multiplex direct amplification reactions underwent library preparation that consisted of fragmentation, ligation of barcoded adapters, size selection and library amplification. Fragmentation was performed using enzymatic reactions developed in the NEBNext® Fast DNA Fragmentation & Library Prep Set for Ion Torrent™ kit. Although fragmentation of the amplicons was successful, the resulting fragments consisted of a large range of sizes and it is thought that a large percentage of the mitochondrial genome represented in the enrichment amplification was not accessible due to over or under digestion. Because the desired size fragment for the Ion Torrent™ PGM™ was 200 base pairs, most of the template source could not used for downstream library preparation. Another fragmentation method that could be better controlled to produce the desired fragment sizes would allow for more of the amplicon

fragments to be included in the library and then sequenced. Sonication, which uses ultrasound waves to shear DNA and may be more easily controlled, may be evaluated in future library preparation testing to produce the desired size fragments for downstream sequencing. Another viable fragmentation option could be a custom restriction endonuclease mixture that targets specific sites in the amplicons to yield a pool of fragments meeting the need to library preparation. Introduction of custom endonuclease fragmentation into the library preparation workflow could also eliminate the need for size selection, resulting in a more efficient method for creating a target enriched library.

Once sequencing was completed, data was uploaded into NextGENe® software for analysis and to align the reads to generate a consesus sequence for each sample. Each sample was assessed to determine the amount of sequencing coverage present over the entire mitochondrial genome. Low coverage was observed in two areas of the genome across all of the samples. The first section was that covered by the 11F/11R amplicon. The second low coverage area was that expected to be covered by the 20F/20R amplicon. Primer binding site polymorphisms that keep the primers from binding efficiently would be a likely explanation for the loss of coverage in these two areas. Because the amplicons used in the multiplex reactions were overlapping, the primer binding sites of 11F, 11R, 20F and 20R were covered by their neighboring amplicons. No polymorphisms in the binding sites of the four primers were observed in the sample consensus sequences.

Another possibility for the low coverage in those two areas is that primer concentration in the primer mixes was too low to produce enough amplicons to be sequenced. Increasing the concentration of these primers in their respective primer mixes may also increase the coverage in the two low coverage areas. Inconsistencies with the multiplex thermal cycling conditions and

the actual melting temperatures (Tm) of the primers could also have resulted in low amplification of the 11 and 20 amplicons. The melting temperatures of the 11F/11R and 20F/20R primer pairs were approximately 55°C and 60°C, respectively. The annealing temperature used in all of the multiplex amplification reactions in this study was 60°C. Adjustment of the 11F/11R primer pair to increase its Tm could result in increased amplification and downstream sequencing coverage of the corresponding mitochondrial region.

Although some loss of mitochondrial genome coverage was observed, over 98% of the genome was covered for each sample. The consensus sequence for each of the pilot samples were analyzed with the MEGA5 software along with the previously-published sequences from Oceanians and Argentinians. The phenogram that depicts genetic distances between each individual indicated the Easter Islanders clustered with most of the Oceanians. The Mapuche samples were in a different cluster along with the Argentinian natives and the remaining Oceanians. These results support the hypothesis that Easter Islanders are more closely genetically-associated with Oceanians than with mainland Chileans (who were represented by the Mapuche in this study). These results also support the occurrence of gene flow between South American natives and Oceanians at some point in history.

It was determined that the mitochondrial regions that were not used in the genetic distance analysis because of sequence coverage loss in the Easter Island and Mapuche samples did not contain differentiating polymorphisms in the previously-published sequences (Figure 12). This suggests that although full mtDNA sequences were not used in the formation of the phenogram, the genetic distance results observed are still accurate. A parsimony tree (Figure 13) depicting the evolutionary relationships between the 43 sequences is concordant with the neighbor joining

tree phenogram (Figure 11) supporting the relationships determined from the pilot set of Eater Islanders and Mapuche natives.

In this study, new multiplex reactions were developed that provided a target enriched mitochondrial genome amplicon pool for library preparation and next generation sequencing. These multiplexes amplified FTA™ based samples that were not able to be amplified with a previously developed multiplex reaction. Additionally, direct amplification from the FTA™ punches with the newly developed primer multiplexes resulted in an amplicon pool ready for library preparation. Adjustments of the 11F/11R and 20F/20R primer pairs, whether it be repacement with new primers, increased concentration in the primer mixes, or modification of amplification parameters, will result in a target enriched library with full coverage of the entire mitochondrial genome. This multiplex assay could then be used for full mitochondrial genome target enrichment for next generation sequencing applications of high quality DNA extracts, mildly degraded samples or FTA™ card preserved samples.

APPENDIX

APPENDIX A

Protocol for Post-Amplification Quantification using the Agilent 2100 Bioanalyzer and Agilent DNA 1000 Kit (Adapted from manufacturer's protocol and UNT Center for Human Identification Procedure Manual)

Preparing the Gel-Dye Mix:

1. Allow DNA dye concentrate and DNA gel matrix to equilibrate to room temperature for 30 minutes.
2. Vortex DNA dye concentrate and add 25 µL of the dye to a DNA gel matrix vial.
3. Vortex solution well and spin down. Transfer to spin filter.
4. Centrifuge at 2240 ± 20% for 15 minutes. Protect from light. Store at 4 °C. Gel-dye mix expires 4 weeks from the date of preparation.

Loading the Gel-Dye Mix:

1. Allow all reagents to equilibrate to room temperature for at least 30 minutes before use. Vortex and briefly centrifuge all reagents.
2. Put a new DNA chip on the chip priming station.
3. Pipette 9.0 µL of gel-dye mix in the well marked with the shaded (G).
4. Make sure the plunger on the chip priming station is positioned at 1 ml and then close the station.
5. Press the plunger until it is held by the clip.
6. Wait for exactly 60 seconds then release the clip.
7. Wait for 5 seconds. Slowly pull back the plunger to the 1 ml position.
8. Open the chip priming station and pipette 9.0 µL of gel-dye mix in the two wells marked with unshaded (G)s.

Loading the Markers:

1. Pipette 5 µL of marker in all 12 sample wells and the ladder well. Do not leave any wells empty.

Loading the Ladder and the Samples:

1. Pipette 1 µL of DNA ladder in the well marked with the ladder symbol.
2. In each of the 12 sample wells pipette 1 µL of sample (used wells) or 1 µL of de-ionized water (unused wells).
3. Put the chip horizontally in the IKA vortex mixer and vortex for 1 minute at 2400 rpm.
4. Run the chip in the Agilent 2100 bioanalyzer within 5 minutes.

APPENDIX B

QIAquick® PCR Purification Kit Protocol (Adapted from Manufacturer's Protocol)

Before Starting:

- This protocol is for the purification of up to 10 μg PCR Products (100 bp to 10 kb in size).
- Add ethanol (96%-100%) to Buffer PE before use (see bottle label for volume).
- All centrifugation steps are carried out at 17,900 x g (13,000 rpm) in a conventional table-top microcentrifuge at room temperature.
- Add 1:250 volume pH indicator I to Buffer PB. The yellow color of the Buffer PB with pH indicator I indicates a pH ≤ 7.5. The absorption of DNA to the membrane is only efficient at pH ≤ 7.5.

Procedure:

1. Add 5 volumes Buffer PB to 1 volume of the PCR reaction and mix, If the color of the mixture is orange or violet, add 10 μL 3M sodium acetate, pH 5.0, and mix. The color of the mixture will turn yellow.
2. Place a QIAquick® column in a provided 2 ml collection tube.
3. To bind DNA, apply the sample to the QIAquick® column and centrifuge for 1 minute. Discard flow-through and place the QIAquick® column back in the same tube.
4. To wash, add 750 μL Buffer PE to the QIAquick® column and centrifuge for 1 minute. Discard flow-through and place the QIAquick® column back in the same tube.
5. Centrifuge the QIAquick® column once more in the provided 2 ml collection tube for 1 minute to remove residual wash buffer.
6. Place each QIAquick® column in a clean 1.5 ml microcentrifuge tube.
7. To elute DNA, add 30 μL elution buffer to the center of the QIAquick® membrane, let the column stand for 1 minute, and then centrifuge for 1 minute.

APPENDIX C

Protocol for Processing of FTA™ Treated Sample Cards for Direct Amplification (Adapted from UNT Health Science Center DNA Identity Laboratory protocol *DNA Extraction from FTA™ Treated Sample Cards*)

Procedure:

1. Using a Harris™ Micro-punch, remove an appropriately sized sample from the center of the stained portion of the FTA™ card for each amplification reaction to be run and eject into a 1.5 ml microcentrifuge tube. (Use the 1.2 mm punch for blood stains and the 2.0 mm punch for buccal stains.)
2. Add 200 µL of FTA™ Purification Reagent to each tube, cap tube, and vortex for 1-2 seconds at low speed.
3. Incubate samples at room temperature for 5 minutes, vortexing after 2.5 minutes.
4. Vortex tubes and remove the Purification Reagent with a transfer pipette.
5. Repeat steps 2-4 two (2) times.
6. Add 200 µL of TE buffer to each tube, cap tube and vortex for 1-2 seconds at low speed.
7. Incubate samples at room temperature for 5 minutes, vortexing after 2.5 minutes.
8. Vortex tubes and remove the TE with a transfer pipette.
9. Repeat steps 6-8 two (2) times.
10. Allow the samples punches to dry in the open microcentrifuge tubes at 56°C for 30 min.
11. Carefully transfer dried sample punches to clean, labeled 200 µL amplification tubes.
12. Amplification master mix can be added directly to the tube and amplification performed. The dry samples can also be stored in the closed tubes at 4 °C until amplification can be performed.

APPENDIX D

Protocol for Direct Amplification of FTA™ Treated Sample Cards (Adapted from Platinum® Multiplex PCR Master Mix Manual)

Preparing the PCR Reaction Mixes:

1. Thaw Primer Mix #1 (PM#1), Primer Mix #2 (PM #2), 2X Platinum® Multiplex PCR Master Mix and Platinum® GC enhancer.
2. Mix Master Mix by inverting tube about 10 times.
3. Place Master Mix in ice block.
4. Mix other reagents by inverting tubes a few times and spinning briefly. Placed tubes on ice block.
5. Prepare the two reaction mixes (one for PM#1 and one for PM #2) as follows:

Reagent	**Volume per reaction (µL)**	**# of samples***	**Total volume (µL)**
2X Master Mix	25 µL	6.6	165 µL
Primer Mix (#1 or #2)	10 µL	6.6	66 µL
GC Enhancer	6 µL	6.6	39.6 µL
De-ionized water	9 µL	6.6	59.4 µL
Total	**50 µL**	6.6	**330 µL**

* # of samples = (samples + negative control + positive control)(1.1 for pipetting error)

6. Invert reaction mix tubes to mix and spin down briefly.
7. Dispense 50 µL of reaction mix to each amplification tube containing the processed FTA™ card punches. (Add 45 µL of reaction mix to positive control and negative control tubes. Then add 5 µL of HL-60 positive control template to positive control tube and 5 µL de-ionized water to negative control tube.)
8. Invert amplification tubes briefly spin down.

Thermal cycling Conditions:

1. Hold 95°C / 2 minutes
2. 35 cycles:
    a. 95°C / 30 seconds
    b. 60°C / 90 seconds
    c. 72°C / 60 seconds
3. Hold 72°C / 10 minutes
4. Hold 4°C / ∞

APPENDIX E

Protocol for the NEBNext® Fast DNA Fragmentation Adapted for Amplicons (Adapted from NEBNext® Fast DNA Fragmentation & Library Prep Set for Ion Torrent™ manual)

Preparing Fragmentation Reaction:

1. Keep reagents and reaction mix on ice when preparing.
2. Mix the following components in amplification tubes (one for each sample) for a total volume of 18.5 µL:

500 ng DNA amplicon pool*	1-15.5 µL
NEBNext® DNA Fragmentation Reaction Buffer	2 µL
NEBNext® $MgCl_2$	1 µL
Sterile water	variable

*Amplicon pool consists of PM#1 and PM#2 amplicons. Concentrations obtained from Agilent electropherograms

3. Vortex for 3 seconds, spin down and place on ice.
4. Vortex the vial of NEBNext® DNA Fragmentation Master Mix for 3 seconds and spin down.
5. Add 1.5 µL of NEBNext® DNA Fragmentation Master Mix to each amplification tube, vortex for 3 seconds and spin down.
6. Incubate in a thermal cycler for 8.5 minutes at 25°C, followed by 10 minutes at 70°C.
7. Pulse spin the amplification tubes and return to ice or keep in -20°C freezer.

APPENDIX F

Protocol for Preparation of Adapter Ligated DNA using NEBNext® Kit Reagents and NEXTflex™ DNA Barcodes (Adapted from NEBNext® Fast DNA Fragmentation & Library Prep Set for Ion Torrent™ manual)

Preparing Adapter Ligated DNA:

1. Add the following to the same amplification tubes containing the fragmentation reaction:

Sterile water	3 µL
NEBNext® T4 DNA Ligase Buffer for Ion Torrent	10 µL
NEBNext® *Bst* DNA Polymerase	1 µL
NEXTflex™ DNA Barcode Adapter*	10 µL
NEXTflex™ DNA P1 Adapter	10 µL
NEBNext® T4 DNA Ligase	6 µL

*A different barcode adapter should be used for each sample (ex: Sample CC-018 with Adapter #1 and Sample CC-023 with Adapter #2)

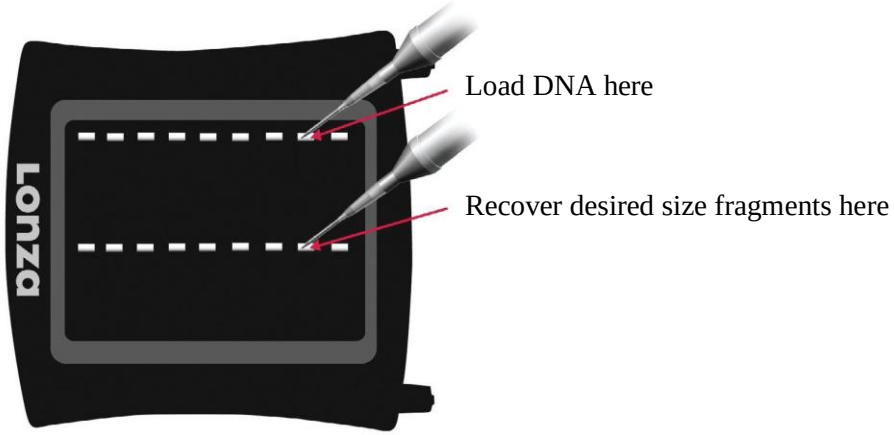
2. Total volume in tubes should be 60 µL.
3. Mix contents by pipetting up and down.
4. Incubate in thermal cycler for 15 minutes at 25°C, followed by 5 minutes at 65°C.
5. Keep tubes on ice or in -20°C freezer.

APPENDIX G

Protocol for Size Selection using the Lonza FlashGel™ System

Size Selection with Lonza Gel:

1. Take FlashGel™ Recovery Cassette (pictured below), rinse wells with sterile water, pour water out and pat areas around wells dry.



Adapted from www.lonza.com

2. Place FlashGel™ Recovery Cassette onto FlashGel™ dock.
3. Add 5 µL FlashGel™ DNA Marker (100 bp - 4 kb) to first (ladder) well
4. Mix 10 µL sample with 2 µL FlashGel™ Loading Dye and add the 12 µL to a well on the gel cassette.
5. Repeat step 3 for each sample.
6. Use power supply to apply 215 volts to gel cassette.
7. Use FlashGel™ Camera to visualize the movement of ladder and samples through the gel cassette.
8. Stop power supply right before desired size fragment enters the recovery well (determined by observing the ladder bands).
9. Remove and discard any liquid in the recovery wells using a pipette.
10. Add 20 µL FlashGel™ Recovery Buffer to each recovery well.
11. Apply 215 volts to gel using power supply.
12. Stop power supply once desired size fragments enter recovery well.
13. Collect the liquid from recovery wells and place in labeled tubes.

APPENDIX H

Protocol for PCR Amplification of Adapter Ligated DNA using NEBNext® Kit Reagents and NEXTflex™ Primer Mix (Adapted from NEBNext® Fast DNA Fragmentation & Library Prep Set for Ion Torrent™ Manual)

PCR Amplification of Adapter Ligated DNA:

1. Mix the following components in sterile amplification tubes (one for each sample) for a total reaction volume of 100 µL:

Adapter Ligated DNA	1-40 µL
NEXTflex™ Primer Mix	10 µL
Sterile water	variable
NEBNext® High-Fidelity 2X PCR Master Mix	50 µL

Thermal cycler Conditions:

1. Hold 98°C / 30 seconds
2. 5 cycles:
    a. 98°C / 10 seconds
    b. 58°C / 30 seconds
    c. 72°C / 30 seconds
3. Hold 72°C / 5 minutes
4. Hold 4°C / ∞

APPENDIX I

Protocol for Amplified Library Clean Up (Adapted from NEBNext® Fast DNA Fragmentation & Library Prep Set for Ion Torrent™ Manual)

Agencourt® AMPure® XP Reagent (Beckman Coulter, Inc., Indianapolis, IN), item number A63882:

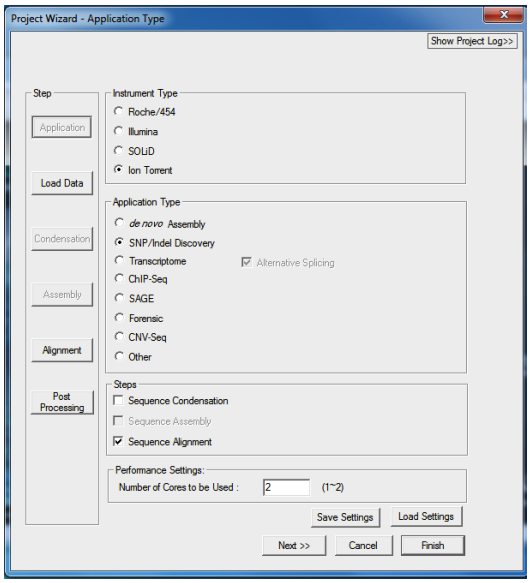
1. Add 1X volume of AMPure® XP Reagent to each PCR reaction and mix by pipetting up and down.
2. Incubate for 5 minutes at room temperature.
3. Spin down the tubes and place in a magnetic rack for 2-3 minutes until the beads have collected to the side of the tube and the solution is clear.
4. Carefully remove and discard the supernatant without disturbing the beads.
5. While the tube is on the magnet, add 500 µL of freshly prepared 80% ethanol.
6. Incubate at room temperature for 30 seconds and carefully remove and discard the supernatant.
7. Repeat steps 5-6.
8. Keep the tubes on the rack, with the caps open, air dry the beads for 5 minutes.
9. Resuspend the beads in 25 µL of 0.1X TE. Mix well on a vortexer and put the tube back on the magnetic rack until the solution is clear.
10. Transfer 20 µL of clear supernatant to a clean, labeled tube.

APPENDIX J

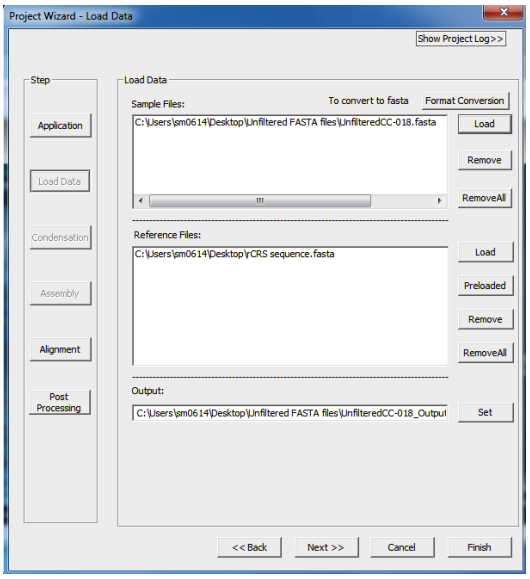
Protocol for read alignment and consensus sequence determination using NextGENe® software

Before opening NextGENe® software: Convert FASTQ files from Ion Torrent PGM™ to FASTA file format using Galaxy Project (http://galaxyproject.org)

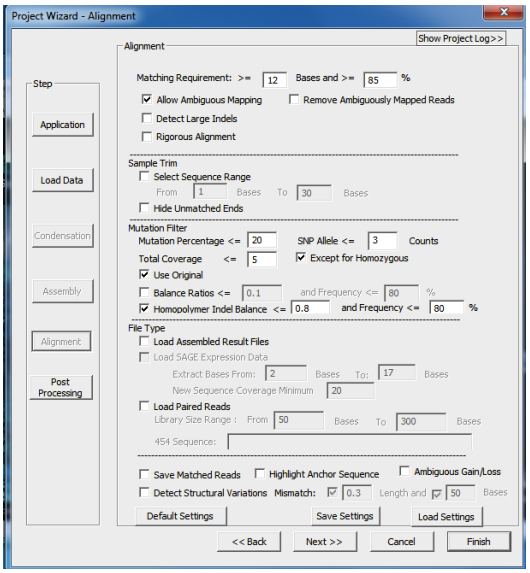
1. Open NextGENe® software program
2. Select File→Open Project Wizard
3. In the Application Type window (see picture at right):
    a. Instrument Type: Select Ion Torrent
    b. Application Type: Select SNP/Indel Discovery
    c. Steps: Check Sequence Alignment
    d. Performance Settings: Number of Cores to be Used: Enter 2
    e. Click Next



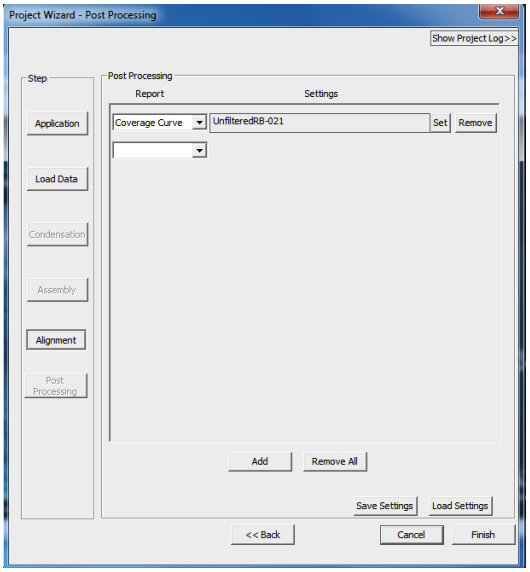
4. In the Load Data Window (see picture at right):
    a. Sample Files: Select fasta file to be aligned to reference file
    b. Reference Files: Select reference file to be aligned to sample file
    c. Output: Select "Set" to create an output folder
    d. Click Next



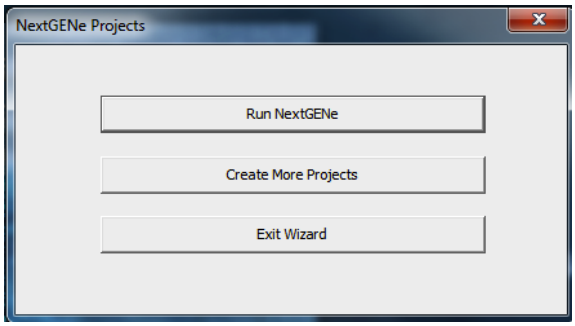
5. In the Alignment Window (see picture at right):
    a. Keep default settings (May change mutation filter to individual preferences)
    b. Click Next



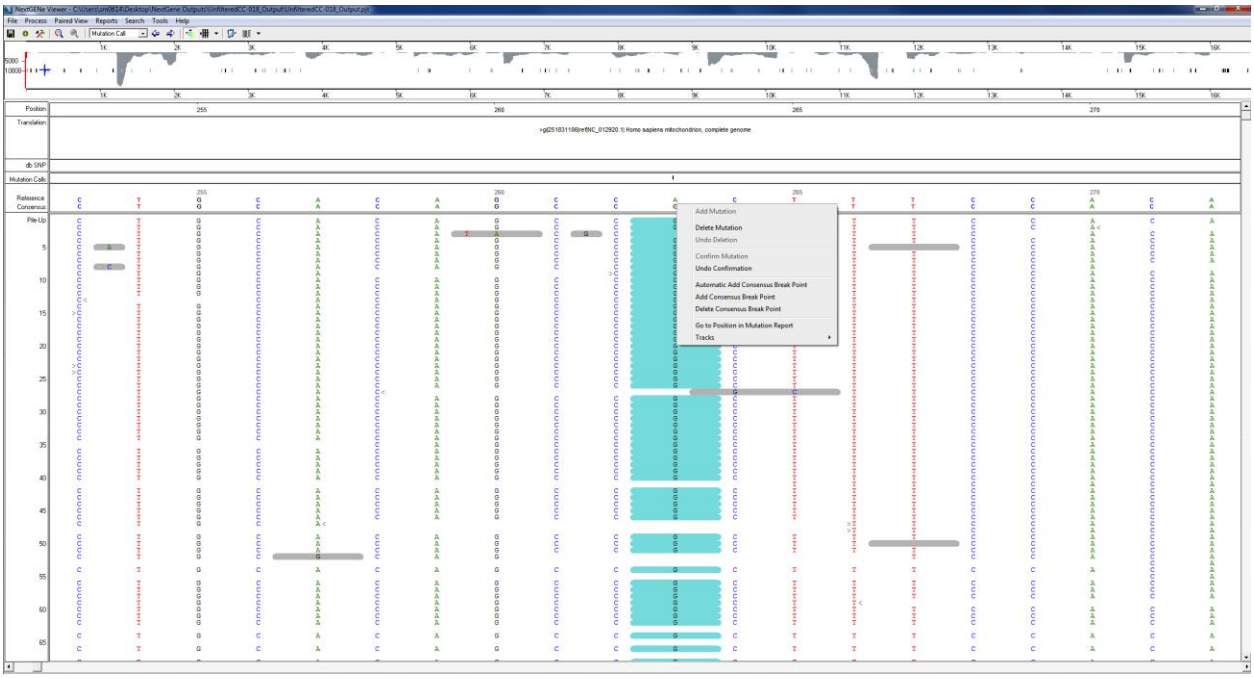
6. In the Post Processing Window (see picture at right):
    a. Select Coverage Curve in drop down menu
        i. Set name for sample coverage curve
    b. Click Finish



7. Click Run NextGENe (see picture at right)



8. NextGENe® output viewer (.pjt file) will automatically open (see picture below).
    a. Coverage map of entire mitochondrial genome shown at top
    b. Reference sequence and computer-determined consensus sample sequence shown below coverage map
    c. Each line below consensus sequence is a sequence read
    d. Grey boxes are computer-determined sequencing errors
    e. Blue boxes are computer determine true variants from the reference sequence
    f. Manual edits to consensus sequence can be made by right-clicking the position to be edited and selecting desired edit to be made.



9. Once final consensus sequence is determined:
    a. Select Reports → Mutation Report Settings→ Save Mutation Report
    b. Select Reports → Mutation Report Settings → Save Consensus Sequence

APPENDIX K

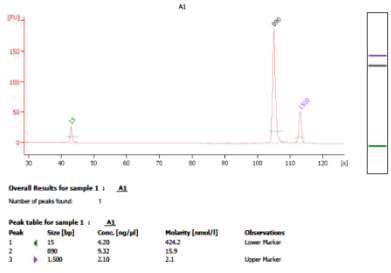
GenBank accession numbers for the previously-published Oceanian and Argentinian full mitochondrial sequences that were used in comparison with the Easter Island and Mapuche samples fully sequenced in this study

GenBank Accession Number	Origin	Reference
AF347002	Papua New Guinea	Ingman *et al.* 2000
AF347003	Papua New Guinea	Ingman *et al.* 2000
AF347004	Papua New Guinea	Ingman *et al.* 2000
AF347005	Papua New Guinea	Ingman *et al.* 2000
AF347007	Samoa	Ingman *et al.* 2000
AJ842744	Taiwan	Trejaut *et al.* 2005
AJ842745	Taiwan	Trejaut *et al.* 2005
AJ842746	Taiwan	Trejaut *et al.* 2005
AJ842750	Taiwan	Trejaut *et al.* 2005
AP012346	Borneo (East Malaysia)	Jinam *et al.* 2012
AP012347	Borneo (East Malaysia)	Jinam *et al.* 2012
AP012348	Borneo (East Malaysia)	Jinam *et al.* 2012
AP012349	Borneo (East Malaysia)	Jinam *et al.* 2012
AP012350	Borneo (East Malaysia)	Jinam *et al.* 2012
AP012351	Borneo (East Malaysia)	Jinam *et al.* 2012
AP012352	Borneo (East Malaysia)	Jinam *et al.* 2012
AY289068	Cook Islands	Ingman and Gyllensten 2003
AY289069	Cook Islands	Ingman and Gyllensten 2003
AY289070	Phillippines	Ingman and Gyllensten 2003
AY289078	Papua New Guinea	Ingman and Gyllensten 2003
AY289079	Papua New Guinea	Ingman and Gyllensten 2003
AY289080	Papua New Guinea	Ingman and Gyllensten 2003
AY289081	Papua New Guinea	Ingman and Gyllensten 2003
AY289093	Samoa	Ingman and Gyllensten 2003
AY289094	Samoa	Ingman and Gyllensten 2003
AY289102	Tonga	Ingman and Gyllensten 2003
GU733718	Phillippines	Gunnarsdottir *et al.* 2011
GU733719	Phillippines	Gunnarsdottir *et al.* 2011
GU733720	Phillippines	Gunnarsdottir *et al.* 2011
GU733722	Phillippines	Gunnarsdottir *et al.* 2011
JN253394	Argentina	Bodner *et al.* 2012
JN253402	Argentina	Bodner *et al.* 2012
JN253418	Argentina	Bodner *et al.* 2012
JN253429	Argentina	Bodner *et al.* 2012
JN253433	Argentina	Bodner *et al.* 2012
JN253421	Argentina	Bodner *et al.* 2012

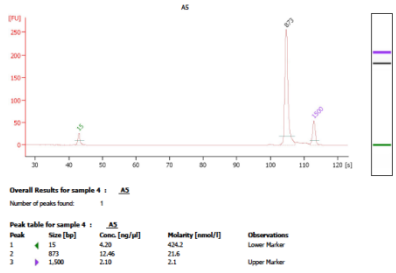
APPENDIX L

Electropherograms of the 24 singleplex reactions for samples CC-012 and TF1

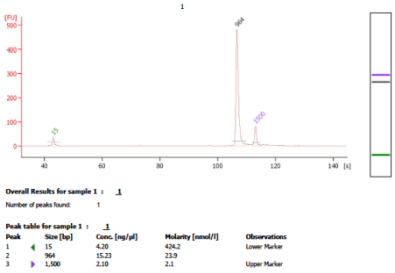
CC-012 1F/1R



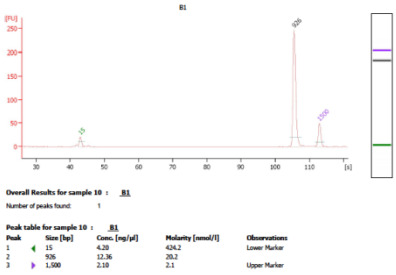
CC-012 2F/2R



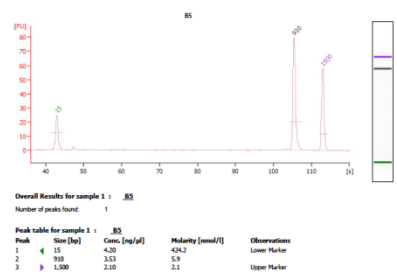
CC-012 3F/3R



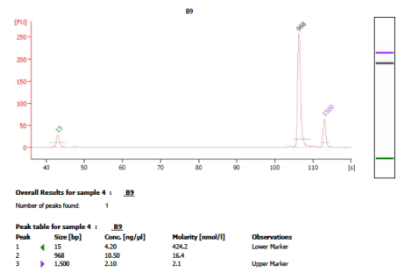
CC-012 4F/4R



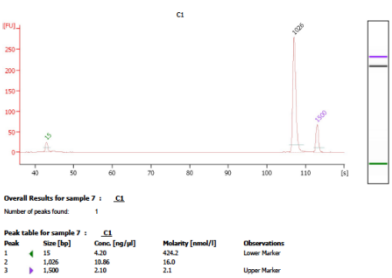
CC-012 5F/5R



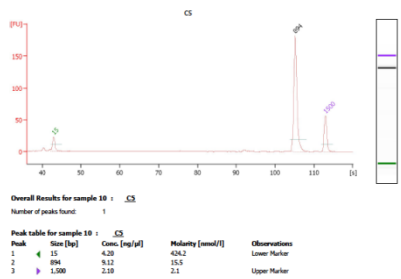
CC-012 6F/6R



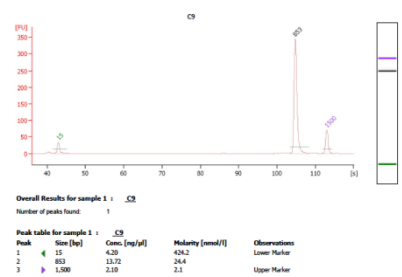
CC-012 7F/7R



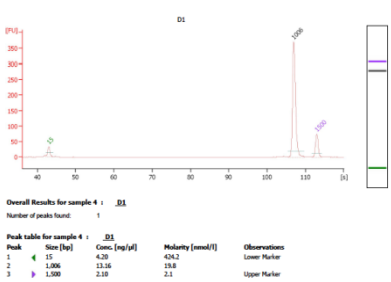
CC-012 8F/8R



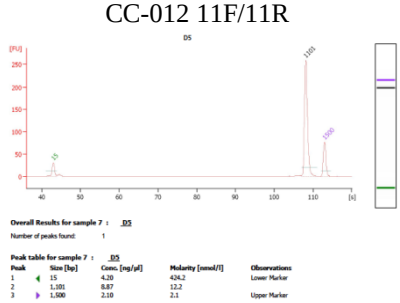
CC-012 9F/9R



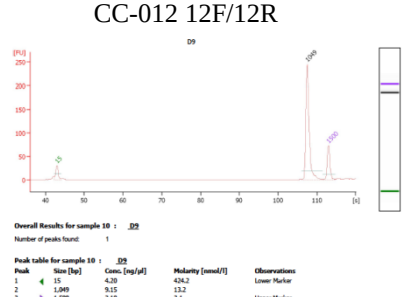
CC-012 10F/10R



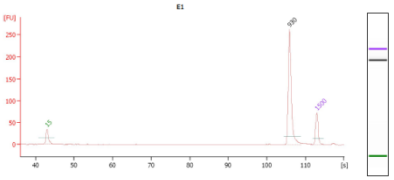
CC-012 11F/11R



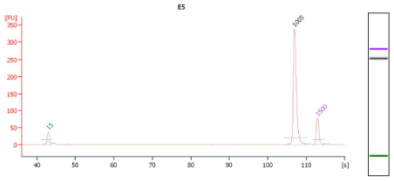
CC-012 12F/12R



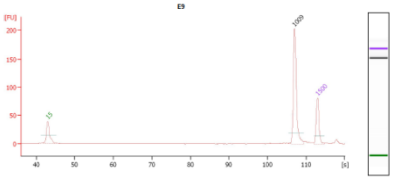
CC-012 13F/13R



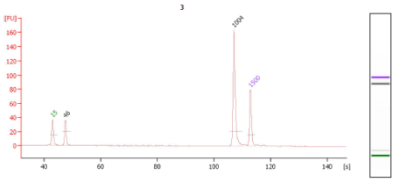
CC-012 14F/14R



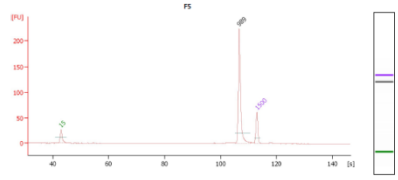
CC-012 15F/15R



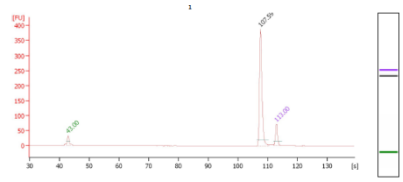
CC-012 16F/16R



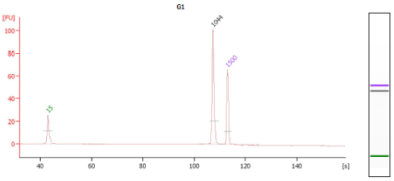
CC-012 17F/17R



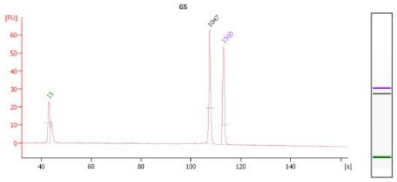
CC-012 F11901/R12876



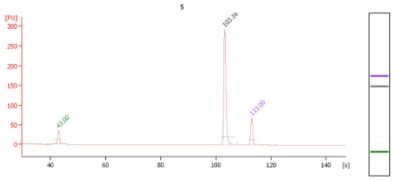
CC-012 19F/19R



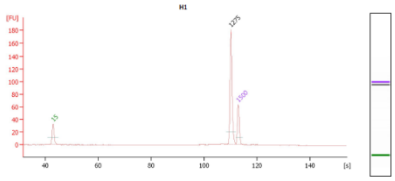
CC-012 20F/20R



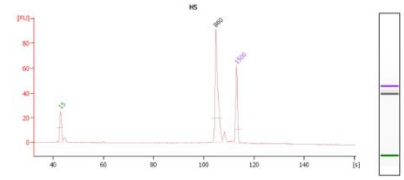
CC-012 F14189/R14924



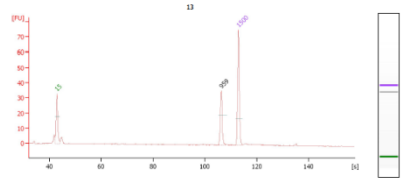
CC-012 22F/22R



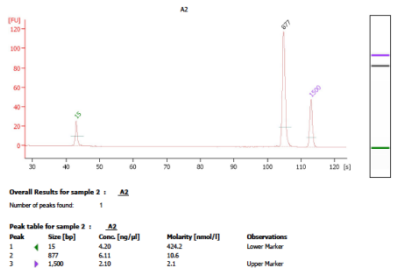
CC-012 23F/23R



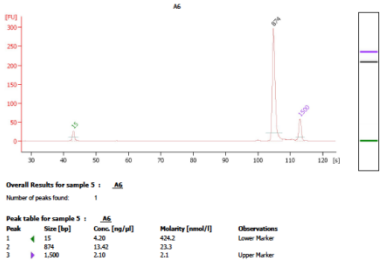
CC-012 16488for/24R



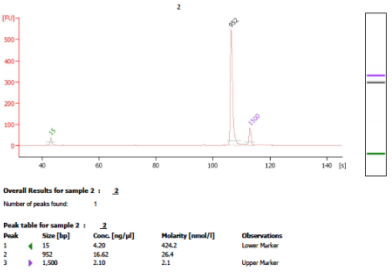
TF1 1F/1R



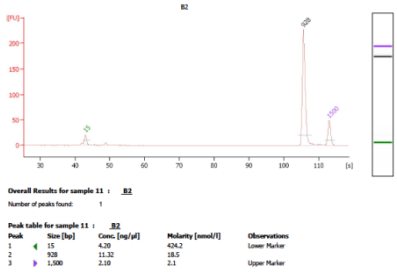
TF1 2F/2R



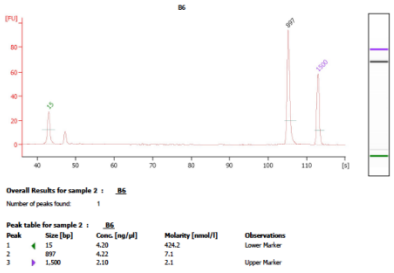
TF1 3F/3R



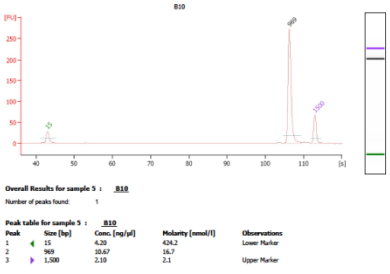
TF1 4F/4R



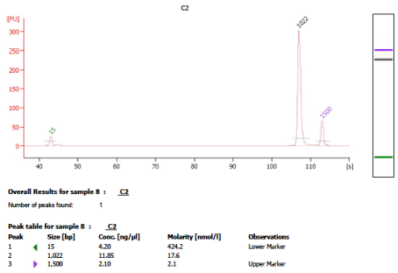
TF1 5F/5R



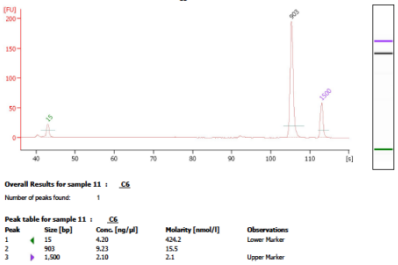
TF1 6F/6R



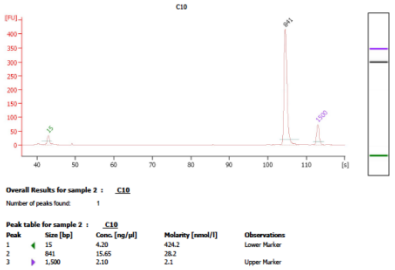
TF1 7F/7R



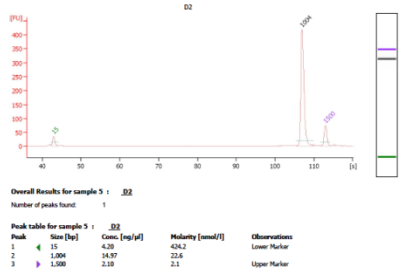
TF1 8F/8R



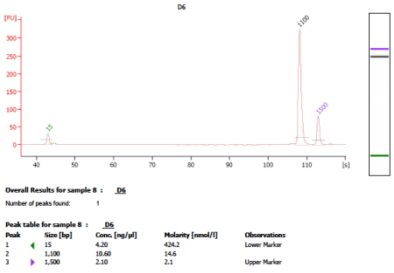
TF1 9F/9R



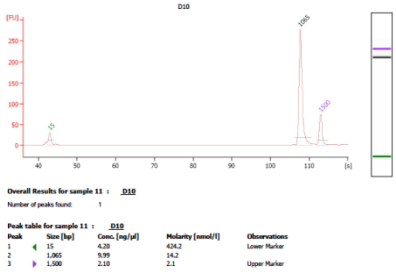
TF1 10F/10R



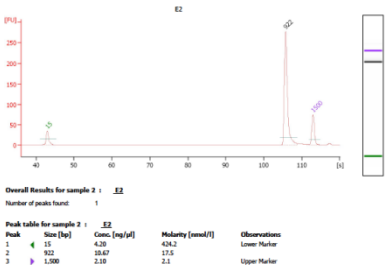
TF1 11F/11R



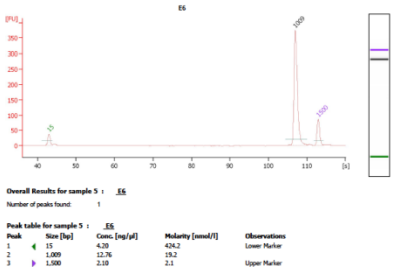
TF1 12F/12R



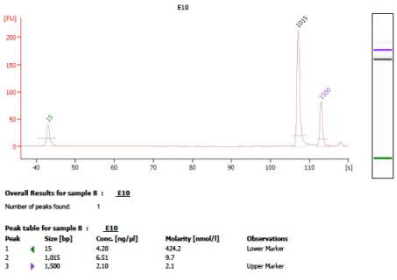
TF1 13F/13R



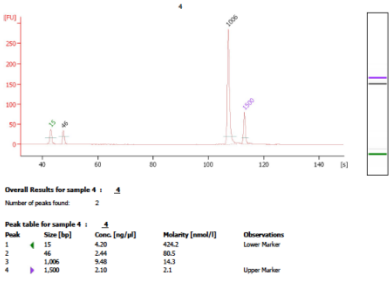
TF1 14F/14R



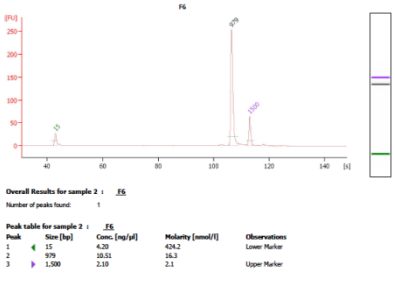
TF1 15F/15R



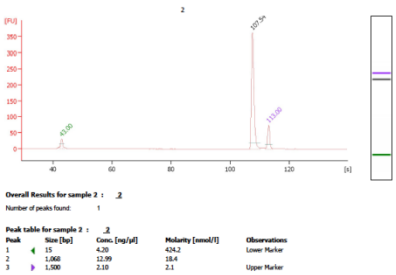
TF1 16F/16R



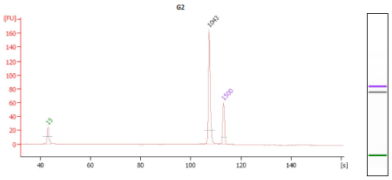
TF1 17F/17R



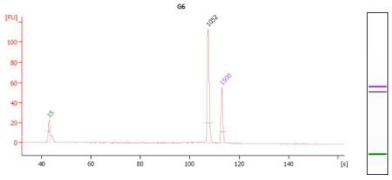
TF1 F11901/R12876



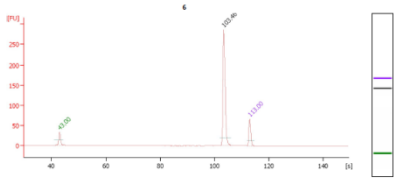
TF1 19F/19R



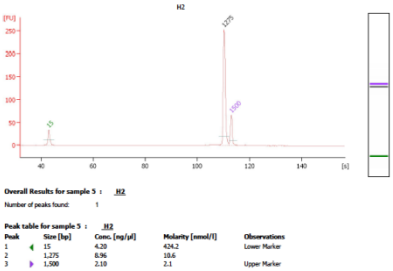
TF1 20F/20R



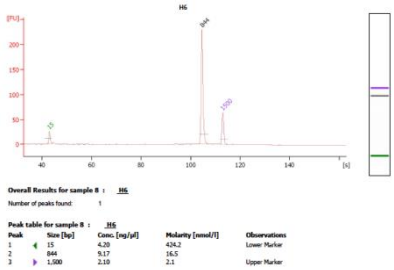
TF1 F14189/R14924



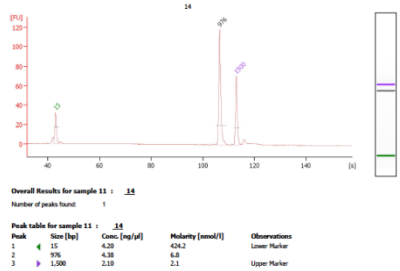
TF1 22F/22R



TF1 23F/23R



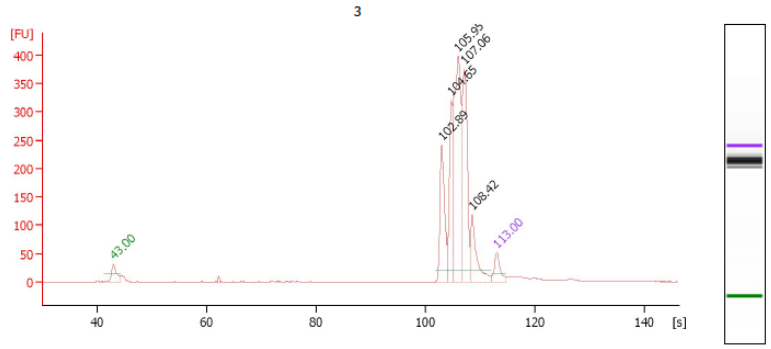
TF1 16488for/24R



APPENDIX M

Electropherograms of extracted samples TF1 and HL-60 amplifed with the two multiplexes developed in this study (Primer Mix #1 and Primer Mix #2)

TF1 Primer Mix #1 - Extracted



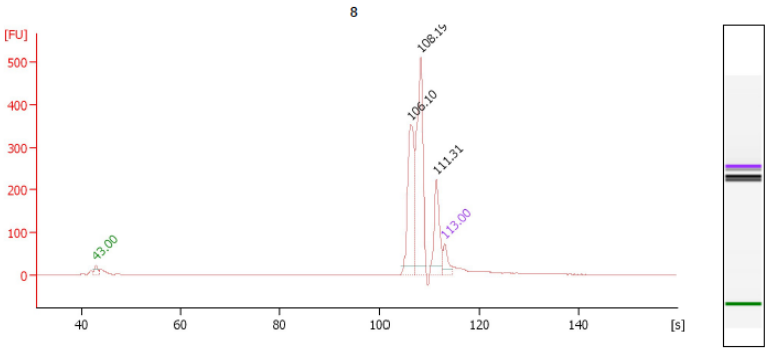
Overall Results for sample 3 : 3

Number of peaks found: 5

Peak table for sample 3 : 3

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	767	9.21	18.2	
3	856	8.48	15.0	
4	946	22.01	35.3	
5	1,027	12.63	18.6	
6	1,135	5.42	7.2	
7	1,500	2.10	2.1	Upper Marker

TF1 Primer Mix #2 - Extracted



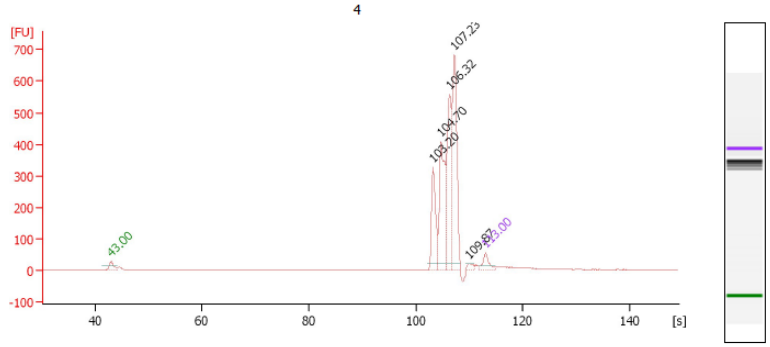
Overall Results for sample 8 : 8

Number of peaks found: 3

Peak table for sample 8 : 8

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	957	15.17	24.0	
3	1,116	20.17	27.4	
4	1,365	7.40	8.2	
5	1,500	2.10	2.1	Upper Marker

HL-60 Primer Mix #1 - Extracted



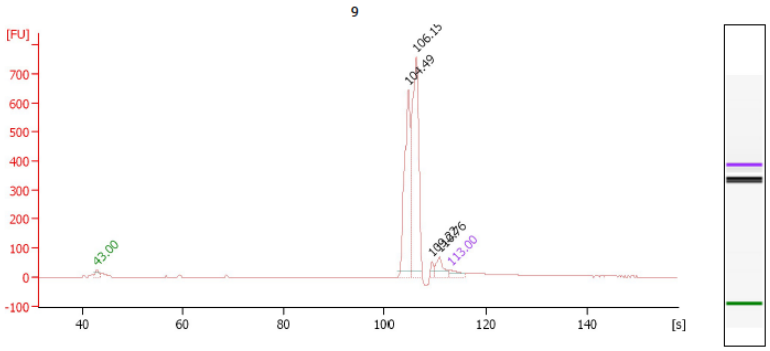
Overall Results for sample 4 : 4

Number of peaks found: 5

Peak table for sample 4 : 4

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	782	9.68	18.7	
3	859	15.04	26.5	
4	972	15.86	24.7	
5	1,040	21.33	31.1	
6	1,251	0.80	1.0	
7	1,500	2.10	2.1	Upper Marker

HL-60 Primer Mix #2 - Extracted



Overall Results for sample 9 : 9

Number of peaks found: 4

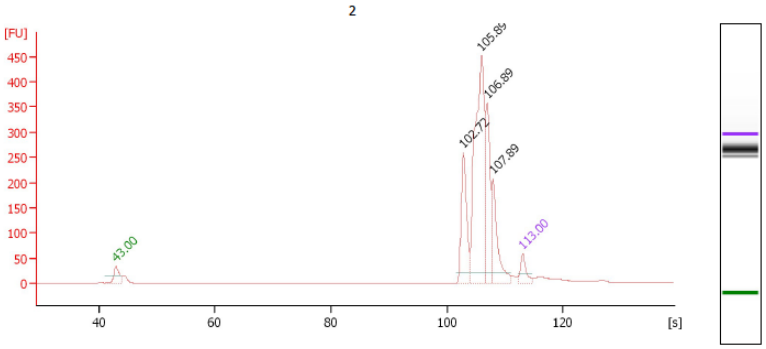
Peak table for sample 9 : 9

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	846	35.23	63.1	
3	960	50.32	79.4	
4	1,199	1.52	1.9	
5	1,322	4.67	5.4	
6	1,500	2.10	2.1	Upper Marker

APPENDIX N

Electropherograms of samples CC-018, CC-023, CC-056, RB-004, RB-009 and RB-021 directly amplified using the two multiplexes developed in this study (Primer Mix #1 and Primer Mix #2)

CC-018 Primer Mix #1 – Direct Amplification



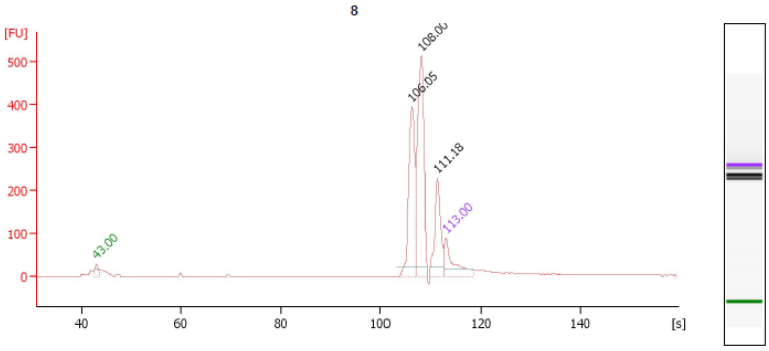
Overall Results for sample 2 : 2

Number of peaks found: 4

Peak table for sample 2 : 2

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	763	9.96	19.8	
3	948	30.96	49.5	
4	1,020	10.71	15.9	
5	1,098	8.04	11.1	
6	1,500	2.10	2.1	Upper Marker

CC-018 Primer Mix #2 – Direct Amplification



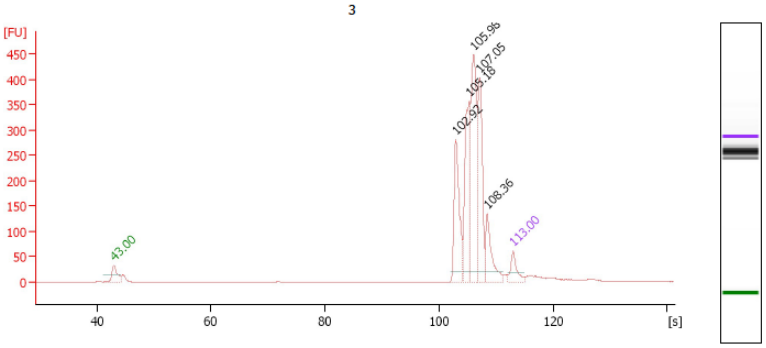
Overall Results for sample 8 : 8

Number of peaks found: 3

Peak table for sample 8 : 8

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	959	7.28	11.5	
3	1,107	9.29	12.7	
4	1,357	3.65	4.1	
5	1,500	2.10	2.1	Upper Marker

CC-023 Primer Mix #1 – Direct Amplification



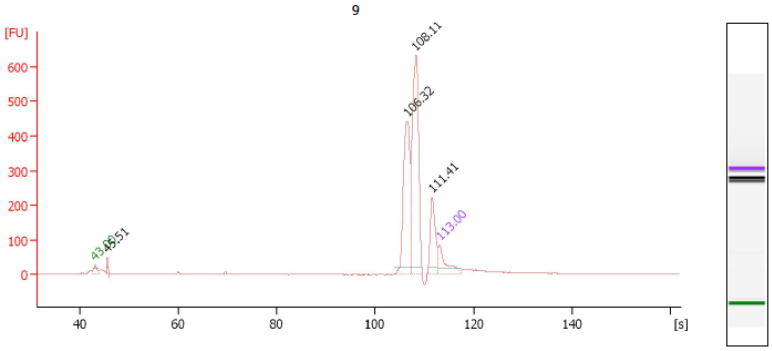
Overall Results for sample 3 : 3

Number of peaks found: 5

Peak table for sample 3 : 3

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	773	9.48	18.6	
3	898	12.46	21.0	
4	954	16.64	26.4	
5	1,033	11.59	17.0	
6	1,135	5.14	6.9	
7	1,500	2.10	2.1	Upper Marker

CC-023 Primer Mix #2 – Direct Amplification



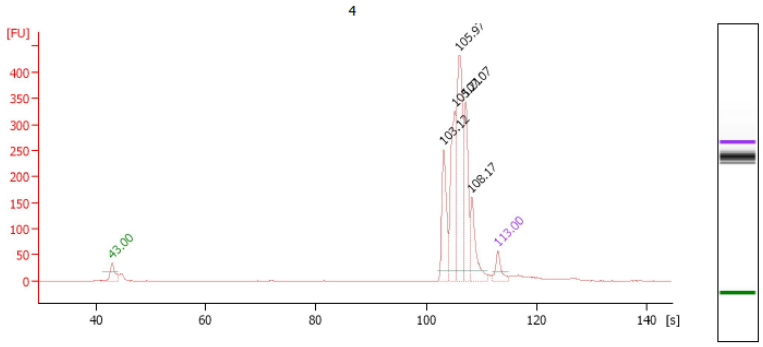
Overall Results for sample 9 : 9

Number of peaks found: 4

Peak table for sample 9 : 9

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	27	0.69	39.4	
3	978	8.68	13.5	
4	1,116	13.44	18.3	
5	1,375	3.71	4.1	
6	1,500	2.10	2.1	Upper Marker

CC-056 Primer Mix #1 – Direct Amplification



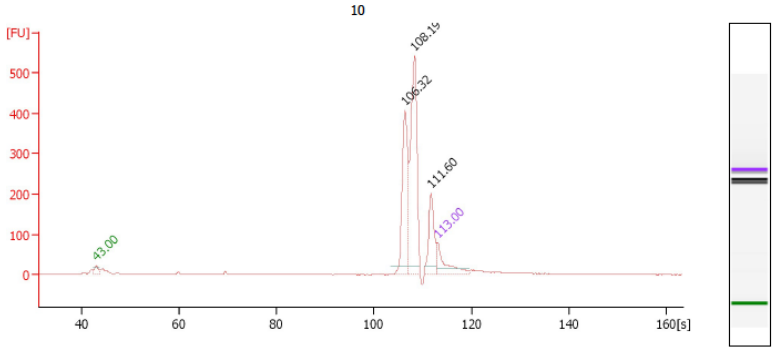
Overall Results for sample 4 : 4

Number of peaks found: 5

Peak table for sample 4 : 4

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	783	8.54	16.5	
3	900	11.96	20.1	
4	953	14.96	23.8	
5	1,034	10.19	14.9	
6	1,121	5.80	7.8	
7	1,500	2.10	2.1	Upper Marker

CC-056 Primer Mix #2 – Direct Amplification



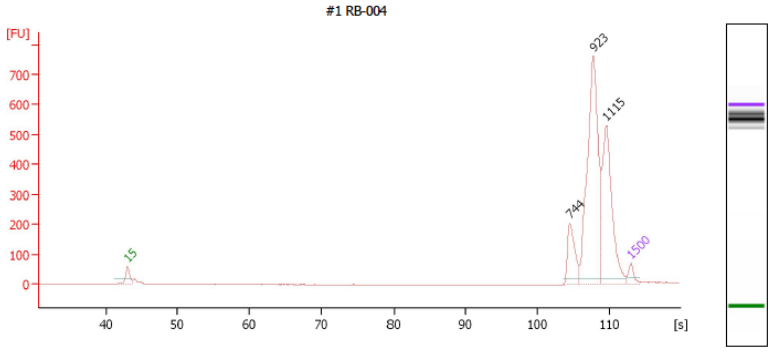
Overall Results for sample 10 : 10

Number of peaks found: 3

Peak table for sample 10 : 10

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	978	7.25	11.2	
3	1,122	10.43	14.1	
4	1,390	3.34	3.6	
5	1,500	2.10	2.1	Upper Marker

RB-004 Primer Mix #1 – Direct Amplification



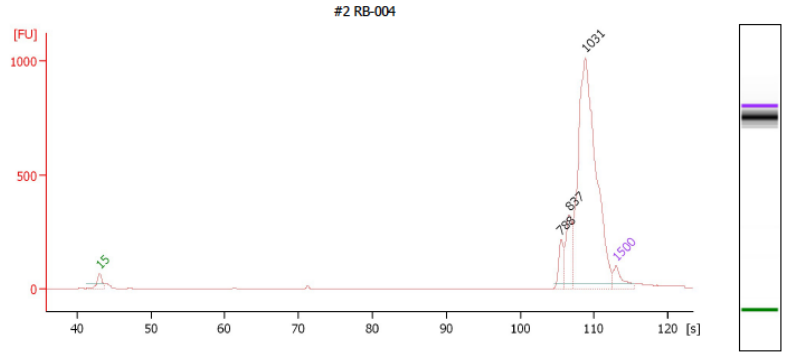
Overall Results for sample 2 : **#1 RB-004**

Number of peaks found: 3

Peak table for sample 2 : **#1 RB-004**

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	744	9.77	19.9	
3	923	52.04	85.4	
4	1,115	31.63	43.0	
5	1,500	2.10	2.1	Upper Marker

RB-004 Primer Mix #2 – Direct Amplification



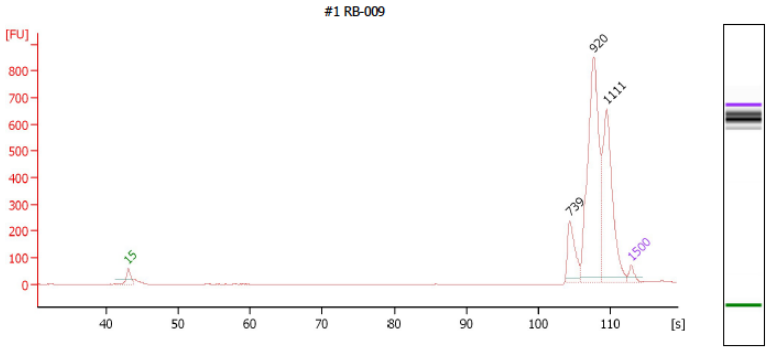
Overall Results for sample 8 : **#2 RB-004**

Number of peaks found: 3

Peak table for sample 8 : **#2 RB-004**

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	788	2.96	5.7	
3	837	5.37	9.7	
4	1,031	47.05	69.2	
5	1,500	2.10	2.1	Upper Marker

RB-009 Primer Mix #1 – Direct Amplification



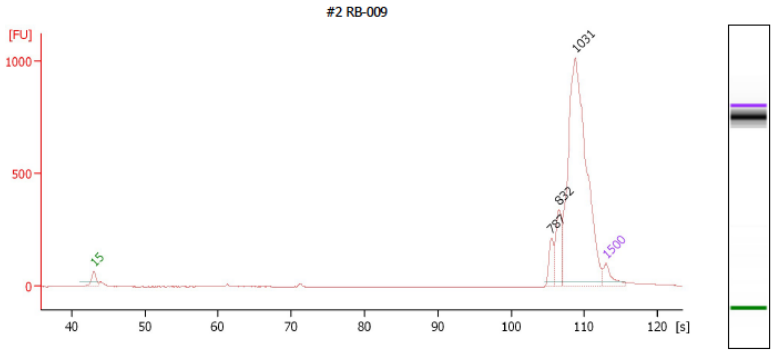
Overall Results for sample 3 : #1 RB-009

Number of peaks found: 3

Peak table for sample 3 : #1 RB-009

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	739	11.53	23.7	
3	920	61.73	101.6	
4	1,111	42.26	57.6	
5	1,500	2.10	2.1	Upper Marker

RB-009 Primer Mix #2 – Direct Amplification



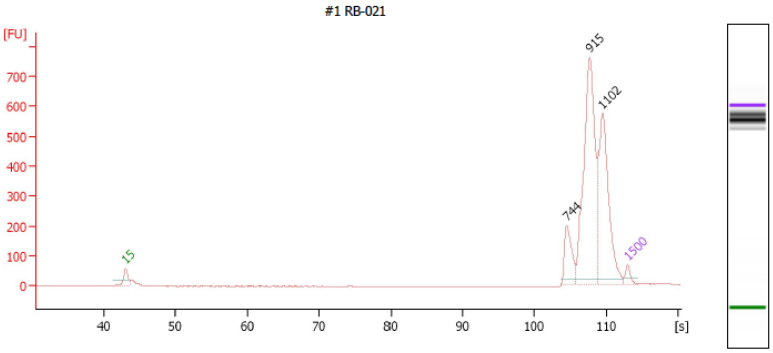
Overall Results for sample 9 : #2 RB-009

Number of peaks found: 3

Peak table for sample 9 : #2 RB-009

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	787	2.66	5.1	
3	832	5.27	9.6	
4	1,031	45.51	66.9	
5	1,500	2.10	2.1	Upper Marker

RB-021 Primer Mix #1 – Direct Amplification



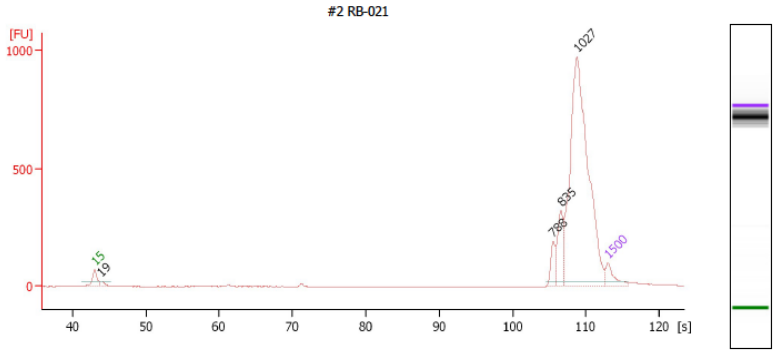
Overall Results for sample 4 : #1 RB-021

Number of peaks found: 3

Peak table for sample 4 : #1 RB-021

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	744	10.40	21.2	
3	915	57.73	95.6	
4	1,102	38.36	52.8	
5	1,500	2.10	2.1	Upper Marker

RB-021 Primer Mix #2 – Direct Amplification



Overall Results for sample 10 : #2 RB-021

Number of peaks found: 4

Peak table for sample 10 : #2 RB-021

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	19	0.90	73.0	
3	788	2.27	4.4	
4	835	4.91	8.9	
5	1,027	42.60	62.9	
6	1,500	2.10	2.1	Upper Marker

REFERENCES

1. Ramos A, Santos C, Alvarez L, Nogués R, Aluja MP. Human mitochondrial DNA complete amplification and sequencing: A new validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification. Electrophoresis. 2009; 30 (9): 1587-93.

2. Butler JM. Forensic DNA typing: Biology, technology, and genetics of STR markers. Academic Press; 2005.

3. Hazkani-Covo E, Zeller RM, Martin W. Molecular poltergeists: Mitochondrial DNA copies (NUMTs) in sequenced nuclear genomes. PLoS Genet. 2010; 6 (2): e1000834.

4. Corser CA, McLenachan PA, Pierson MJ, Harrison GA, Penny D. The Q2 mitochondrial haplogroup in Oceania. PloS One. 2012; 7 (12): e52022.

5. Lopez JV, Culver M, Stephens JC, Johnson WE, O'Brien SJ. Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. Mol Biol Evol. 1997; 14 (3): 277-86.

6. Li M, Schroeder R, Ko A, Stoneking M. Fidelity of capture-enrichment for mtDNA genome sequencing: Influence of NUMTs. Nucleic Acids Res. 2012.

7. Yao Y, Kong Q, Salas A, Bandelt H. Pseudomitochondrial genome haunts disease studies. J Med Genet. 2008; 45 (12): 769-72.

8. Lie B, Dupuy B, Spurkland A, Fernández-Viña M, Hagelberg E, Thorsby E. Molecular genetic studies of natives on Easter Island: Evidence of an early European and Amerindian contribution to the Polynesian gene pool. Tissue Antigens. 2007; 69 (1): 10-8.

9. Thorsby E, Flåm S, Woldseth B, Dupuy B, Sanchez-Mazas A, Fernandez-Vina M. Further evidence of an Amerindian contribution to the Polynesian gene pool on Easter Island. Tissue Antigens. 2009; 73 (6): 582-5.

10. Ghiani ME, Moral P, Mitchell RJ, Hernández M, García-Moro C, Vona G. Y-chromosome-specific STR haplotype data on the Rapanui population (Easter Island). Human Biol. 2006; 78 (5): 565-78.

11. González-Pérez E, Esteban E, Via M, García-Moro C, Hernández M, Moral P. Genetic change in the Polynesian population of Easter Island: Evidence from Alu insertion polymorphisms. Ann Hum Genet. 2006; 70 (6): 829-40.

12. Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, et al. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? The Am J Hum Genet. 2001; 69 (6): 1348-56.

13. Rieder MJ, Taylor SL, Tobe VO, Nickerson DA. Automating the identification of DNA variations using quality-based fluorescence re-sequencing: Analysis of the human mitochondrial genome. Nucleic Acids Res. 1998; 26 (4): 967-73.

14. Levin BC, Hancock D, Holland K, Cheng H, Richie K. Human mitochondrial DNA--amplification and sequencing standard reference materials--SRM 2392 and SRM 2392-I. US Department of Commerce, National Institute of Standards and Technology; 2003.

15. Pennisi E. Semiconductors inspire new sequencing technologies. Science. 2010; 327 (5970): 1190.

16. Life Technologies. Ion Library Quantification Kit User Guide. 2011; Catalog Number 4468802.

17. Life Technologies. Ion OneTouch™ 200 Template Kit v2 DL User Guide. 2012; Catalog Number 4480285.

18. Life Technologies. Ion PGM™ Sequencing 200 Kit v2 User Guide. 2012; Catalog Number 4482006.

19. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011; 28 (10): 2731-9.

20. Ingman M, Kaessmann H, Pääbo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. Nature. 2000; 408 (6813): 708-13.

21. Ingman M, Gyllensten U. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. Genome Res. 2003; 13 (7): 1600-6.

22. Gunnarsdóttir ED, Li M, Bauchet M, Finstermeier K, Stoneking M. High-throughput sequencing of complete human mtDNA genomes from the Philippines. Genome Res. 2011; 21 (1): 1-11.

23. Jinam TA, Hong L, Phipps ME, Stoneking M, Ameen M, Edo J, et al. Evolutionary history of continental Southeast Asians: "Early train" hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. Mol Biol Evol. 2012; 29 (11): 3513-27.

24. Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, Hsu CJ, et al. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. PLoS Biol. 2005; 3 (8): e247.

25. Bodner M, Perego UA, Huber G, Fendt L, Rock AW, Zimmermann B, et al. Rapid coastal spread of first Americans: Novel insights from South America's Southern Cone mitochondrial genomes. Genome Res. 2012; 22 (5): 811-20.

26. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of the mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993; 10 (3): 512-26.

27. Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M. The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. Am J Hum Genet. 1995; 57 (6): 1463-75.