A22 Entire mtGenome Sequencing: A Strategy for High-Quality Samples

Elizabeth A. Lyons, MFS, Kimberly Sturk-Andreaggi, MFS, Jodi A. Irwin, PhD, and Rebecca S. Just, MFS, 1413 Research Boulevard, Rockville, MD 20850*

After attending this presentation, attendees will gain an understanding of considerations particular to entire mtGenome sequencing and insight into the development and implementation of an automated mitochondrial genome (mtGenome) protocol for population and reference samples.

This presentation will impact the forensic science community by providing an amplification and sequencing strategy designed to produce complete forward and reverse sequence coverage over the entire mtGenome with minimal manual reprocessing.

There has been a steady rise in the number of entire mitochondrial genome (mtGenome) haplotypes generated for medical genetic, phylogenetic and population studies, as well as for forensic applications in recent years. Although the number of complete human mitochondrial DNA (mtDNA) sequences in GenBank now exceeds 6500, many of these sequences contain errors and few meet the criteria for use as forensic reference data. Given this, along with new techniques that will simplify access to mtDNA coding region data in forensic specimens, the creation of mtGenome reference databases appropriate for forensic use is needed.[1] However, the development of complete mtGenome haplotypes is labor intensive, expensive and fraught with opportunities for human error. An optimized, automated process is thus essential for high volume generation of mtGenome reference data that meet forensic standards.

The Armed Forces DNA Identification Laboratory (AFDIL) has generated more than 500 complete mtDNA sequences using a 12- amplicon, 108-reaction Sanger sequencing strategy over the past ten years.[2,3] Though the process results in error-free data, the protocol, originally developed more than a decade ago,[4] frequently requires extensive manual reprocessing to produce complete haplotypes. Therefore the goal of this work was to develop an automated amplification and sequencing strategy for high-quality (non-degraded) samples that would routinely produce data sufficient to cover the entire mtGenome.

Suitable placement for amplification and sequencing primers was assessed using published mtGenome substitution rate data and haplogroup-specific polymorphism information.[5,6] A web-based tool was then used for amplification and sequencing primer design. Candidate amplification primer pairs were subjected to varying PCR conditions to determine optimal thermal cycling parameters, and a capillary-based detection instrument was used to gauge amplification efficiency. Enzymatic, column, and bead-based purification protocols were evaluated to balance efficacy, cost, and automation. Sequencing primers were subsequently tested to assess sequence data quality, defined by the degree of "background" sequence and the range of high-resolution data.

An eight-amplicon mtGenome strategy that allows eleven samples to be processed on each 96-well amplification plate was designed. Amplification set-up was performed on a robotic instrument to ensure correct sample placement, and an optimal PCR extension time has been implemented. Amplification success was assessed using an automated capillary electrophoresis system that requires no manual pipetting. Amplicons, which range in size from 1,197 to 2,544 base pairs, were

sequenced in 16 reactions each for a total of 128 sequencing reactions per sample. Alternating forward and reverse sequence primer placement ensured redundant sequence coverage across the entire mtGenome. All post-PCR pipetting steps (post-amplification purification, sequencing set- up, post-sequencing purification and sequence detection set-up) were performed robotically, and sequence detection was performed on a 48-capillary electrophoresis instrument.

This optimized, highly automated protocol reduced cost, hands-on laboratory time, and opportunities for human error by substantially decreasing the number of manual production steps and the extent of sample reprocessing necessary to construct complete mtGenome haplotypes. The high-throughput strategy will facilitate regular generation of high-quality entire mtDNA profiles for forensic reference databases and other applications which would benefit from error-free mtDNA data.

References:

1. J.A. Irwin et al., "mtGenome reference population databases and the future of forensic mtDNA analysis," *Forensic Sci. Intl. Genet*. (2010) doi:10.1016/j.fsigen.2010.02.008.
2. M.D. Coble et al., "Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of the forensic testing in Caucasians," *Intl. J. Legal Med.* (2004) 118:137-146.
3. R.S. Just et al., "Complete mitochondrial genome sequences for 265 African-American and U.S. Hispanic individuals," *Forensic Sci. Intl. Genet.* (2008) 2:e45-e48.
4. Levin et al., "A human mitochondrial DNA standard reference material for quality control in forensic identification, medical diagnosis and mutation detection," *Genomics* (1999) 55:135-146.
5. P. Soares et al., "Correcting for purifying selection: An improved human mitochondrial molecular clock," *Am. J. Hum. Genet.* (2009) 84:740-759.

 ** Presenting Author*

6. M. Van Oven, M. Kayser, "Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation," *Hum. Mutat.* (2009) 30:e386-e394.

Mitochondrial DNA, Sequencing, Protocol

 ** Presenting Author*