RATES AND MATERNAL TRANSMISSION OF mtDNA HETEROPLASMY USING A NEXT GENERATION SEQUENCING APPROACH

Jennifer McElhoe[1], Mitchell Holland[1], Kateryna Makova[2], Marcia Shu-Wei Su[2], Anton Nekrutenko[3], Boris Jaramillo[3], Christine Baker[4], Seth Faith[5], and Brian Young[5]
Forensic Science Program[1], Biology Department[2] and Biochemistry & Molecular Biology Department[3], The Pennsylvania State University
Battelle Charlottesville Operation[4]
Battelle Columbus Operation[5]

Sequencing of mitochondrial DNA (mtDNA) recovered from damaged, degraded, or limited biological samples is a powerful tool used by forensic scientists. For decades, capillary electrophoresis based Sanger sequencing has been the gold standard in DNA sequencing, but this technology is inherently hampered by limitations in speed, throughput, resolution, and associated costs. The advent of next generation sequencing (NGS) technologies has revolutionized genomic studies with greater throughput and reduced cost. The deep sequencing data produced by NGS platforms increases the resolution of mtDNA heteroplasmic sites, and in turn, will impact the interpretation of mtDNA comparisons between evidentiary and known samples encountered in forensic casework.

In this study we assess the occurrence and rate of low level variants in mtDNA collected from 174 individuals, including more than 50 maternally related pairs. All samples were sequenced on Illumina's MiSeq platform using paired-end, 250 bp reads. Our approach of whole mtDNA genome amplification using two overlapping (~8,500 bp) amplicons and sample multiplexing using Illumina's Nextera XT library preparation kit, consistently produced runs with an average of 85% of the base calls having a quality score >30 on the Phred scale (1 in 1,000 probability of an incorrect base call). Each individual run on the MiSeq multiplexed 12 samples and produced a read coverage of at least 1,000 across the entire mtDNA genome, with much of the coverage above 10,000 reads. This level of coverage allows for detection of minor heteroplasmic variants (below 1%), and provides the basis for an assessment of coverage requirements for the reporting of these low-level components.

This presentation will focus on:

- Methods development for generating whole mtDNA genome sequence data on the MiSeq instrument
- Analysis parameters for identification and reporting of low-level heteroplasmic variants
- Assessment of coverage requirements for reporting low-level variants