MODELING HETEROPLASMIC RATE ESTIMATES FOR ANALYSIS OF MITOCHONDRIAL DNA

M.M. Holland[1], P. Wilton[2], W. Parson[1,3], and J.A. McElhoe[1]
[1]The Pennsylvania State University
[2]University of California
[3]Institute of Legal Medicine, Medical University of Innsbruck

Massively parallel sequencing (MPS) technologies present an opportunity to extend forensic mitochondrial (mt) DNA typing capabilities beyond the current standard. Now that mixed sites of mtDNA sequence can be routinely resolved, forensic laboratories are able to report heteroplasmy on a routine basis. Statistical models will be needed to determine the relative frequency of observing an mtDNA haplotype, when including the presence of a heteroplasmic site. Based on 1,300 mtDNA control region (CR) sequences from buccal swabs collected from four major population groups (European NIJ 2014-DN-BX-K022, and African, Asian, and Latino NIJ 2016-DN-BX-0171), the rates of heteroplasmy on an individual and nucleotide position (np) basis were determined, and a statistical model was established for calculating the frequency of haplotype/heteroplasmy matches. Sequencing was performed on amplified products targeting the entire CR using a Nextera XT library preparation and paired-end reads on an Illumina MiSeq. Using a minor allele frequency threshold of ≥2%, the overall rate of heteroplasmy was 37.7%, with mixed sites observed at 12.5% of the nps across the CR. The majority of heteroplasmic positions (75.2%) had frequencies of 2-10%, and most individuals (76.6%) had one observation of heteroplasmy. Statistical analysis evaluating the possible linkage between haplotype and heteroplasmy included broad (X^2 test) and phylogenetic association testing (maximum-likelihood phylogeny estimator RAxML), and an estimation of the heteroplasmy occurrence rate per haplogroup (Bayesian approach). P-values were negatively correlated with the number of observations of heteroplasmy for both the X^2 test (Spearman's rho=-0.330, p=0.0037) and phylogenetic test (Spearman's rho=-0.236, p=0.0405), suggesting that in many cases the small sample size causes a lack of statistical association. Thus we concluded the different rate of heteroplasmy observed between major haplotype groups is likely to occur at many nps. The estimation of heteroplasmy frequency per haplogroup at each position for each major haplogroup showed a mean rate of 0.152 (±0.134) with a minimum of 0.031 and a maximum of 0.83. Identifying a robust statistical model will be a key feature of applying MPS techniques for the analysis of mtDNA sequence in forensic cases. This will have a positive impact on laboratories currently performing mtDNA analysis, and could be the impetus for other laboratories to bring mtDNA typing on line as discrimination potential is improved through the reporting of heteroplasmy.