WEIGHT ESTIMATES AND STATISTICAL EVALUATION OF THE INDEPENDENCE OF HETEROPLASMY AND HAPLOTYPE BASED ON RATE DATA FROM A EUROPEAN POPULATION GROUP

Jennifer A. McElhoe, Ph.D. and Mitchell M. Holland, Ph.D., Forensic Science Program, Department of Biochemistry & Molecular Biology, The Pennsylvania State University

The presence of mitochondrial (mt) DNA heteroplasmy in a forensic case can significantly enhance the weight of a resulting match between evidence and reference sources, highlighting a long-held desire of the forensic community to move towards the routine reporting of heteroplasmic sequences. A likelihood ratio (LR) can be used in forensic science to convey weight estimates of an mtDNA sequence match. While the LR based on the haplotype is relatively straightforward, the calculation becomes more complicated with the presence of heteroplasmy, as the calculation of a LR must take both the haplotype frequency and the probability of observing heteroplasmy at any one position along the mtDNA control region (CR) into account. If the haplotype and heteroplasmy are independent, statistics that employ a frequency for each can be multiplied together, while dependency will require combined haplotype and heteroplasmy frequencies to be addressed. Providing LR calculations will require a better understanding of the rates of heteroplasmy and a thorough evaluation of possible linkage using a robust statistical model. In this study we established haplotype and heteroplasmy rates for 550 CR sequences of European descent and evaluated independence by applying statistical significance tests to the most commonly observed haplotypes and heteroplasmic sites using Fisher's exact testing. To increase sample sizes, analyses compared haplotypes combined into haplogroups with both individual heteroplasmic observations (local approach) and groups of heteroplasmic sites based on position in the CR region (global approach). Our analysis produced mixed results, with the global analysis indicating dependence while the local approach suggested that the majority of individual heteroplasmic observations were independent with five sites (nucleotide position/haplogroup; 16093/K; 16189/U; 215/U; 185/H; 185/J) possessing possible dependence. These results convey that a site by site assessment of independence would be the best method to establish the significance of a match involving heteroplasmy. It should be noted that these results are preliminary, and while we believe this statistical analysis is a valid approach for assessing possible linkage, the rarity of heteroplasmic observations presents a challenge for the simulated approximation produced by any statistical analysis and will benefit from additional collection of data. We will present our findings and apply them to case scenarios, including the identification of Nicholas Romanov, the last Russian Tsar.