TOOLS AND CONSIDERATIONS FOR MITOCHONDRIAL HAPLOGROUP ASSIGNMENT

Kimberly Sturk-Andreaggi[1,2,3], Nicole Huber[4], Walther Parson[4,5], Marie Allen[3], Charla Marshall[1,2]
[1]ARP Sciences, LLC; Rockville, MD
[2]Armed Forces DNA Identification Laboratory, Armed Forces Medical Examiner System; Dover, DE
[3]Department of Immunology, Genetics and Pathology, Uppsala University; Uppsala, Sweden
[4]Institute of Legal Medicine, Innsbruck Medical University; Innsbruck, Austria
[5]Forensic Science Program, The Pennsylvania State University; University Park, Pennsylvania

Forensic guidelines recommend the use of a sample's matrilineal phylogeny, denoted by its haplogroup, as a valuable QC measure for mtDNA analysis. Multiple variants not consistent with the haplogroup may indicate an issue with haplotype quality. The haplogroup is also necessary to guide phylogenetic nomenclature, which impacts database searching. Phylotree is a comprehensive mtDNA phylogenetic tree that details the diagnostic polymorphisms of all currently recognized mtDNA haplogroups. Manual haplogroup determination can be achieved using Phylotree but is complicated and time-consuming, especially with small regions sequenced in forensics. Automated tools such as SAM2 (incorporated into EMPOP) and HaploGrep2 are routinely used by the forensic community and allow easy determination of mtDNA haplogroups. Additionally, the AQME Mitochondrial Haplogrouper was developed by a forensic laboratory for use during sequence data analysis with the CLC Genomics Workbench. However, the accuracy of the mtDNA haplogroup assignment is dependent on both the tool's algorithm and region of the mitogenome used to make the assignment.

Haplogrouping accuracy was evaluated by estimating the haplogroup for 92 diverse entire mitogenome haplotypes with the three tools as well as manually using Phylotree. SAM2 haplogroups were 100% identical to the manual assignments. HaploGrep2 and AQME each had one (1%) haplogroup that differed from the manual/SAM2 assignment, which were less refined by 1 or 2 nodes. Accuracy of the tools was further investigated by performing haplogroup estimations based on three smaller mtDNA regions: the control region (CR), hypervariable segments (HVS) 1 and 2 and HVS1 alone. The smaller ranges resulted in a loss of resolution regardless of the tool used or haplotype. Of note, haplogroups from different clades (relative to the mitogenome predictions) were assigned based on HVS1 data for eight samples with SAM2, 11 with HaploGrep2, and 14 with AQME. A major advantage of SAM2 is that the most recent common ancestor (MRCA) is automatically reported when more than one haplogroup was identified, whereas the other tools required an analyst to determine the MRCA haplogroup. Overall, SAM2 predictions were more accurate with 96% (176) of the 184 CR and HVS1-2 haplogroups consistent with the mitogenome assignments (compared to 94% of the HaploGrep2 predictions and 91% of the AQME predictions). Features and application may also be considerations in determining the practical utility of a haplogroup tool within a forensic laboratory.

The opinions or assertions presented are the private views of the author(s).