EVALUATION OF GeneMarker® HTS FOR IMPROVED ALIGNMENT OF mtDNA MPS DATA, HAPLOTYPE DETERMINATION, AND HETEROPLASMY ASSESSMENT

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

Massively parallel sequencing (MPS) technologies present an opportunity to extend forensic mitochondrial (mt) DNA typing capabilities beyond the current standard. In our lab we have analyzed data sets containing more than 700 mtDNA sequences, illustrating that the increased resolution provided by MPS produces a higher rate of resolvable heteroplasmy. However, the process of post sequencing evaluation has proven to be challenging, and a rate-limiting step. Existing software has not allowed for the proper alignment of mtDNA sequence data, resulting in the inaccurate assessment of haplotypes and variant frequencies, and requiring considerable manual analysis. The regions of sequence that are typically difficult to align are homopolymeric stretches, isolated patterns of SNPs (single nucleotide polymorphisms), and INDELs (insertions/deletions). A software solution (GeneMarker® HTS, SoftGenetics, Inc.), which uses a customizable motif-based alignment strategy, was developed and evaluated to address these limitations and provide a user-friendly interface for forensic practitioners and others interested in mtDNA analysis of MPS data. The HTS software was developed to address the following; 1) alignment to a circular version of the mtgenome so that data properly spans the transition point in the mtgenome numbering system, 2) recognition of SNP-associated motifs and INDELs (insertions/deletions) consistent with phylogenetic and forensic considerations, 3) identification of heteroplasmic sequences, and 4) export of reports that address forensic considerations and allow for seamless import into tertiary analysis tools such as EMPOP (www.empop.org). Sequence data from 500 individuals, with various alignment asymmetries and levels of heteroplasmy, were used to experimentally assess the software. Accuracy in producing mtDNA haplotypes, the ability to correctly identify low-level heteroplasmic sequence variants, and the user-based features of the software were all evaluated. Our results suggest that HTS can be used with MPS data to accurately generate phylogenetically correct haplotypes and heteroplasmic positions. The software is easy to use, offers numerous user-defined parameters for filtering the data, and provides multiple options for viewing and navigating through the data.