A146 Expert System Rules for Mitochondrial DNA Sequence Analysis

Nicole R. Phillips, MS, University of North Texas Health Sciences Center, Department of Forensic and Investigative Genetics, 3500 Camp Bowie Boulevard, Room 310, Fort Worth, TX 76107; and Rhonda K. Roby, PhD, University of North Texas, Center for Human ID, 3500 Camp Bowie Boulevard, Room 310, Fort Worth, TX 76107*

After attending this presentation, attendees will understand the need for increased automation and standardization in forensic mtDNA sequence analysis and how software programs can function as an expert system to facilitate objective, high-throughput sequence data analysis.

This presentation will impact the forensic science community by highlighting a need for increased consistency and throughput of mtDNA data analysis and demonstrating current advances in expert system development that will potentially offer a solution.

Mitochondrial DNA (mtDNA) analysis has proven to be an invaluable tool for victim identification in mass disasters, missing persons programs, and criminal casework. The University of North Texas Center for Human Identification, primarily funded by the National Institute of Justice (NIJ) for the Missing Persons Program, uses advanced DNA technologies to process unidentified human remains and the family reference samples from biological relatives for both nuclear DNA and mtDNA. Since most missing person cases rely heavily on mtDNA testing of skeletal remains, mtDNA testing of appropriate family reference samples is necessary for recommending familial associations. The resulting DNA profiles are uploaded to the Missing Persons Index database. In this database, mtDNA and nuclear DNA profiles from the unidentified remains can be searched against the biological family reference profiles and associations are recommended through kinship analysis testing. There are thousands of missing persons cases reported each year, with more than 14,000 unidentified human skeletal remains stored in medical examiners' and coroners' offices nationwide. These numbers alone demonstrate the throughput needs for DNA processing.

Sample processing methods are continually improving with advances in liquid handling robotics, chemistry, and instrumentation. Consequently, the rate of data generation exceeds that of data analysis, review, and reporting, hence creating a bottleneck in the final upload of data and reporting. The bottleneck in STR data analysis was specifically addressed by NIJ's Expert System Testbed Project, which emphasized the use of expert system technologies for STR data analysis. The resulting

implementation of such software programs has augmented backlog reductions across the country. However, there currently is not a software program or package that is capable of automating the more complex process of mtDNA sequence data analysis. This presentation discusses the unique challenges of automating mtDNA sequence data analysis and the solutions that are in development at the University of North Texas Center for Human Identification.

Three primary challenges in analyzing mtDNA data are: (1) managing the volume of sequence data generated, which greatly exceeds that of STR analysis; (2) base calling, which is more subjective than allele calling; and, (3) assigning haplotypes, which by nature is subject to inconsistency between individual analysts as well as laboratories. Initial software package developments have been made to address each of the three challenges while providing a workflow interface which facilitates laboratory processing and data management. Expert system rules have been designed to automate analysis at the trace level, the contig level, and the haplotype level, including new advances in eFAST™ Software (University of North Texas Health Science Center, Ft. Worth, TX), the front-end program designed to expedite trace quality assessment. Unique trace-level expert system rules such as High Baseline (HB), Mixture (M), Length Heteroplasmy (LH), Homopolymeric Stretch (HPS), and Low Signal (LS) have been programmed and tested to increase the consistency and throughput of trace quality assessment. For example, High Baseline alerts the analyst if a trace has underlying signal that exceeds a user- defined threshold. The Mixture rule fires to alert the analyst that a trace contains a number of mixed bases that exceeds the user-defined threshold. The Length Heteroplasmy and Homopolymeric Stretch rules fire when the indicated anomaly is encountered in the sequence trace and eFAST Software will then adjust the Contiguous Read Length expectation. Such rules not only increase the accuracy and resolution of the filter metrics, but also give the analyst valuable information at a glance to guide further sample processing. These developments are the first steps towards providing the forensic mtDNA testing community with a comprehensive expert system that seamlessly takes the analyst from data collection through haplotype reporting in order to increase the accuracy and efficiency of mtDNA data analysis and management.

mtDNA, Expert System, eFAST

 ** Presenting Author*