



B18 The Design and Compilation of a Comprehensive U.S. Y-STR Haplotype Reference Database

Paulina N. Berdos, BS and Erin K. Hanson, MS, National Center for Forensic Science, University of Central Florida, Department of Chemistry, PO Box 162367, Orlando, FL 32816-2367; Stacey Smith, MS, South Dakota DCI Forensic Laboratory, 3500 E Highway 34, Pierre, SD 57501; Jack Ballantyne, PhD, National Center for Forensic Science, University of Central Florida, Department of Chemistry, PO Box 162367, Orlando, FL 32816-2367*

After attending this presentation, the forensic community will learn the design, establishment and statistical analysis of a comprehensive online Y-STR Database.

This project would impact the forensic community because it will provide reliable estimates of Y-STR haplotype frequencies and provide a statistical estimate of the significance of a match.

The establishment of a U.S. National Y-STR reference database from a variety of geographically and ethnically diverse populations is essential to facilitate the generation of reliable estimates of Y-STR haplotype frequencies. Such multi-locus haplotype frequencies are required to provide a statistical estimate of the significance of a match. A U.S. YSTR Haplotype Reference Database has been created by the International Forensic Y-User Group and is maintained by the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. However, this database has been limited to a set of 9 core Y-STRs that limits its operational usefulness, particularly in light of the development of Y-STR multiplexes consisting of additional Y-STR loci. Y-STR loci, unlike traditional STR markers, are not independent of one another and are co-inherited as extended haplotypes of linked markers. The estimation of the frequency of occurrence of a particular haplotype therefore necessitates the use of a counting method which means that the significance of many matches are dependent upon the size of the database.

The NCFS database initially comprised data generated in the laboratory based upon a 19 Y-STR locus extended haplotype. The loci include DYS19, DYS385 (a) and (b), DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS425, DYS434, DYS437, DYS438, DYS439, Y-GATA C4, Y-GATA A7.1, Y-GATA A7.2 and Y-GATA H4. Data was obtained from various Caucasian, African American, Hispanic, Native American and Asian populations. A comprehensive evaluation of the data was carried out to determine the allelic distribution and the gene diversity at each locus as well as the multilocus discriminatory potential. Allelic ladders for the 19 Y-STR markers have been constructed to aid in accurate genotype determinations. Since there are a number of common 19-locus haplotypes additional loci have been included in the database. These loci include DYS426, DYS436, DYS441, DYS450, DYS462, YAP (Alu insertion), Y-GATA-A10, DYS446, DYS435, DYS442, DYS443, DYS444, DYS445, DYS447, DYS448, DYS449, DYS452, DYS453, DYS454, DYS455, DYS456, DYS458, DYS463, DYS464, DYS468, DYS484, DYS522, DYS527, DYS531, DYS557 and DYS588. Although, some data exists for some of these loci in U.S. populations, it is not readily accessible to the crime laboratory community and usually does not contain individual multilocus haplotype data.

A key component of our strategy is to allow for the continuous updating of haplotype data using the same samples. This ensures that as new markers are developed, the same samples would be re-typed, and a new extended haplotype developed. Thus, any laboratory needing haplotype data for any combination of Y-STR markers would be served. The aid of geographically diverse crime laboratories has been enlisted to obtain the necessary samples. In exchange for the samples the crime laboratories benefit by obtaining a custom built no-cost local Y-STR database.

Our results to date will be presented, and information on community access will be provided.

Y-SR Database, Haplotype Reference Database, Populations Studies