



B51 The Design and Compilation of a National Y-STR Haplotype Reference Database

Paulina Berdos, BS, Ashley Hall, MS, and Erin Hanson, BS, Department of Chemistry, University of Central Florida, P.O. Box 162366, Orlando, FL; and John Ballantyne, MS, PhD, National Center for Forensic Science, P.O. Box 162367, Orlando, FL*

The goal of this presentation is to inform the forensic community on the design, establishment and statistical analysis of a comprehensive on-line Y-STR database, the aim of which is to type all potentially useful Y-STR markers in a variety of geographically and ethnically diverse populations.

The establishment of a U.S. National Y-STR reference database is essential to facilitate the generation of reliable estimates of Y-STR haplotype frequencies. Such haplotype frequencies are required to provide a statistical estimate of the significance of a match. A U.S. Y-STR Haplotype Reference Database has been created by the International Forensic Y-User Group and is maintained by the Max Plank Institute for Evolutionary Anthropology, Leipzig, Germany. However, this database is limited to a set of 9 core Y-STRs which limits its operational usefulness, particularly in light of the development of Y-STR multiplexes consisting of over 40 different 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 is entirely dependent upon the size of the database.

The database records initially comprise data generated in the laboratory based upon a 19 Y-STR locus extended haplotype. The loci tested 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 compiled from various Caucasian, African American, and Hispanic populations. Although some (unpublished) data exists for some of these loci in U.S. populations, it is not readily accessible to the crime laboratory community and usually does not contain 'extended' haplotype data due to the technological restraints of the systems employed by the investigators. A key component of the 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 was 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.

The database is in the process of being extended to include 42 loci, and the results of that effort will be described.

Y-STR Database, Haplotype Reference Database, Statistical Analysis