A26 Y - STR Haplotype Database Comparison for Colorado Residents

Sylvia Thurmond, MFS, Bonnie Mountain, MA, and Susan Berdine, MS, Denver Police Department, FBIO/DNA Unit, 1331 Cherokee Street, Denver, CO 80204*

After attending this presentation, attendees will understand the importance of evaluating, supporting, and defending the selection of a publicly available Y-STR haplotype database for estimating haplotype frequencies.

This presentation will benefit the forensic community by providing an example of how to evaluate and determine the appropriate Y-STR haplotype database through the use of haplotype data collected from residents of a specific geographical region (in this case, Colorado).

The counting method is a simple and conservative method used to estimate the frequency of a Y-STR haplotype. This method relies on searching a specific Y-STR haplotype in a database of reference Y-STR haplotypes from unrelated, random individuals of self-described population groups (e.g., African American, Hispanic, and Caucasian) and utilizing the number of times that haplotype occurs in the database to estimate the upper bound frequency estimate using a 95% confidence interval. The width of the confidence interval is inversely related to the size of the database and may be used to estimate how often the specific haplotype would be expected to be observed in any given database (e.g., the entire US population). A number of web-accessible Y-STR haplotype databases are available to the forensic community (e.g., YHRD, Applied Biosystems, and Promega were publicly available at the time of this study). These databases vary in terms of size and searchable Y-STR loci. Due to the different database choices available to the forensic community, and to respond to challenges regarding the use of “pooled” versus “regional” databases, this study was designed to determine if the upper bound frequency estimate for Y-STR haplotypes generated from Colorado residents varied significantly when different haplotype databases were utilized.

At the Denver Police Department Crime Laboratory, Y-STR haplotypes were generated using the AmpfℓSTR® Yfiler® Y-STR PCR Amplification kit from 38 individuals who are employees of the Denver Police Department Crime Laboratory and are residents of Colorado. Complete 17-locus profiles were generated from all 38 samples, but for searching and comparison purposes, the 11-locus US Haplotype (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, YS385a/b, DYS438, and DYS439) was searched in the three available databases. The upper bound frequency estimate was determined for each searchable population group and the most conservative value (i.e., maximum upper bound frequency) was compared between the three databases for each haplotype. For 35 of the 38 Colorado samples, the Applied Biosystems database (N=3,561) provided the highest maximum upper bound frequency estimate and for 37 of the 38 Colorado samples, the YHRD database (N=22,999) provided the lowest maximum upper bound frequency estimate. These results are consistent with the view that the larger the database becomes, the more precise the frequency estimate is due to the width of the confidence interval decreasing. Therefore, due to the size of the database, the Applied Biosystems Yfiler® Haplotype database provided the most conservative upper bound frequency estimate for the majority of the Y-STR haplotypes searched. Also, it would be expected that the discriminatory power of the Applied Biosystems database is more powerful than this study reflects and a lower upper bound frequency estimate would be obtained if all 17 Y-STR loci were searched in the database, rather than limiting the search to the 11-locus US Haplotype.

Since the completion of this study, the US Y-STR database was released and includes haplotype data from five different sources (i.e., National Center for Forensic Science, ReliaGene, Promega, Applied Biosystems and the University of Arizona) pooled into a single searchable database. Current work is underway to search this database with the same 38 Colorado 11-locus haplotypes to determine if this larger database is consistent with the previous results of this study. This study is expected to benefit the relevant forensic community by providing an example of how to evaluate and determine the appropriate Y-STR haplotype database through the use of haplotype data collected from residents of a specific geographical region (in this case, Colorado).

Y-STR, Database, Comparison

* *Presenting Author*