A Y-STR MIXTURE FREQUENCY ESTIMATOR

Kevin MacMillan[1], M.S., Lisa Gefrides[1], M.S., Cindi Klein[1], M.S., Lyn Fatolitis[2,3], B.S., Jack Ballantyne[2,3], Ph.D., and Roger Kahn[1], Ph.D.
[1]*Harris County Medical Examiner's Office, Houston, TX*
[2]*University of Central Florida, Department of Chemistry, Orlando, FL*
[3]*National Center for Forensic Science, Orlando, FL*

In crime scene mixtures, forensic DNA typing using autosomal short tandem repeats (STRs) may not show a male component due to an excess of female DNA. Y-STR testing can be employed to visualize the male contributor to the mixture. Y-STR testing targets the male Y chromosome to generate a male-only profile. The resulting profile is termed a haplotype.

It is always desirable to estimate the frequency of a profile, or in this case a haplotype, in order to give statistical weight to a match between a crime scene sample and a known reference sample. To estimate the frequency of occurrence of a multi-locus, autosomal forensic DNA test, the genotype frequencies from each locus are multiplied together since each locus is inherited independently. STR loci on the Y chromosome are not inherited independently. Instead, they are transmitted as a set without re-assortment. Barring mutation, the combination of alleles on the Y chromosome will be identical to that of a male's biological father as well as other male paternal relatives. To estimate the frequency of a Y-STR haplotype, the counting method is used: the frequency is equal to the number of times a particular Y-STR haplotype appears in a database for a given population divided by the total number of haplotypes in the database.

To obtain reliable estimates a large Y-STR database is required. Several groups have pooled their databases to form the US Y-STR database (http://usystrdatabase.org/) managed by the National Center for Forensic Science (NCFS) in conjunction with the University of Central Florida. As of December 31, 2007 the US Y-STR database contained 13,906 profiles with complete, 11-locus SWGDAM-core haplotypes. The database allows profile frequencies to be estimated by users over the Internet.

It would be advantageous to provide a frequency estimate when Y-STR mixtures are obtained but currently there is no way to search the US Y-STR Database with mixture data. We have developed a template using Microsoft Office EXCEL that allows the input of all alleles found in a Y-STR mixture to determine how many haplotypes from the US Y-STR database could be included in the mixture. This template will accommodate both single source and mixture samples with 10 alleles per locus, or more in scalable fashion, reporting the frequency estimates as a total and by race. Both full and partial profiles may be used for searching. The template may be used with any Y-STR typing system.