Y-STR DATABASE

The US Y-STR Database

By Lyn Fatolitis and Jack Ballantyne
National Center for Forensic Science
University of Central Florida, Orlando, Florida, USA

The US Y-STR Database is a comprehensive and searchable Y-STR reference database containing more than 13,000 11- to 17-locus Y-STR haplotypes divided into five forensically relevant ancestries.

The National Center for Forensic Science (NCFS), a program of the National Institute of Justice hosted by the University of Central Florida, in conjunction with the Y-STR Consortium created at the American Academy of Forensic Science meeting in 2006 has created a large comprehensive Y-STR reference database of more than 13,000 haplotypes, which is now available online at: **www.usystrdatabase.org** (Figure 1). The US Y-STR Database, a searchable listing of 11- to 17-locus Y-STR haplotypes, was developed by combining data from NCFS with online databases maintained by the University of Arizona, Applied Biosystems, Inc., ReliaGene, Inc., and Promega Corporation (Figure 2).

The database provides tools to obtain Y-STR haplotype frequencies needed to calculate matching or paternity probabilities with confidence intervals. Other features include the ability to simultaneously upload multiple haplotypes for searches directly from Genotyper® and GeneMapper® text files, the ability to include or exclude sampled populations, and a report-style printout of the results. Samples are divided into five forensically relevant ancestries: African-American, Asian, Caucasian, Hispanic and Native American.

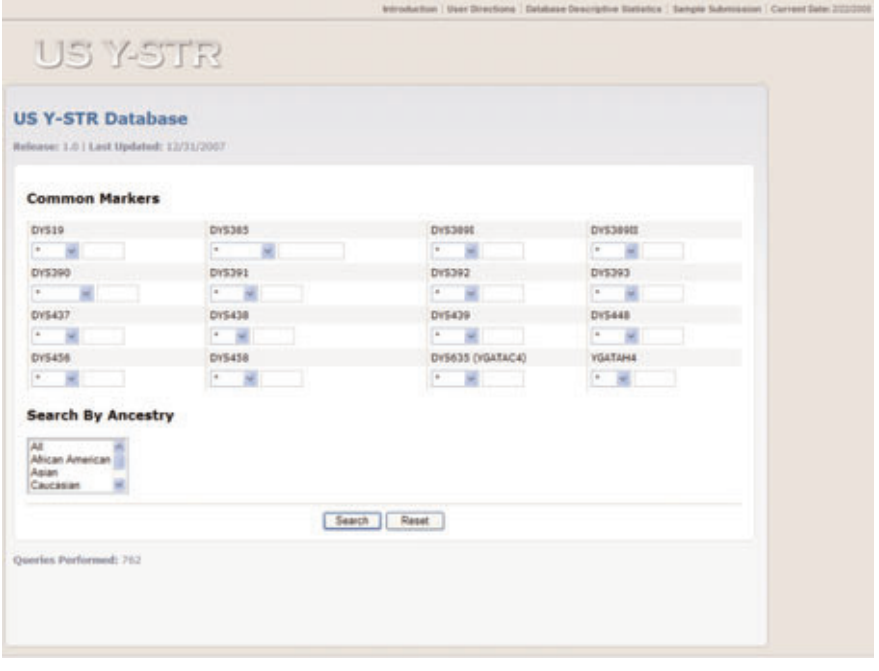

Figure 1. The US Y-STR Database interface.

Y-STR DATABASE

The goal of the database is to expand continuously the number of individuals (N) for each ancestral group and geographical location. NCFS is currently creating quality assurance procedures to govern the suitability of data solicited from diverse sources for inclusion in the database, including a proficiency testing procedure for labs who wish to contribute Y-STR haplotypes in the future, ensuring that they can correctly genotype samples. Information about submitting Y-STR data to expand the database will soon be accessible from the database web site.

It is important to note that a number of individual samples were shared among the contributing data sets. All duplicate samples were removed to ensure that each sample in the consolidated database is from a unique individual. Any population group that did not contain at least 50 samples was also removed. These data reconciliation and reorganization steps have resulted in the consolidated US Y-STR Database having slightly different sample numbers than those found in the curated databases currently maintained by the individual contributing institutions. This population database is intended for use in estimating haplotype population frequencies for forensic casework purposes. All donors are anonymous, and original electropherograms do not exist in a curated fashion. All submitting entities are solely responsible for their data. In the event that details of a certain population sample are requested via the judicial process, the request will be redirected to the collaborating scientists and their institutions.

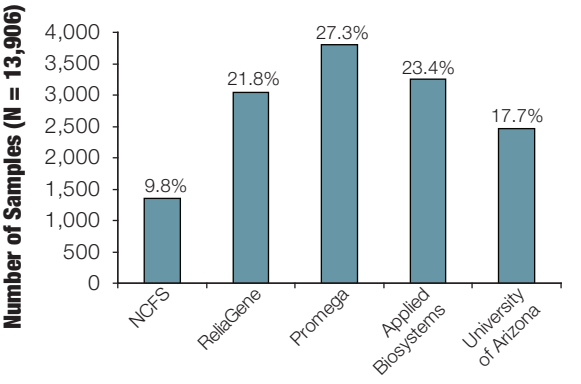

Figure 2. Data contributors to the US Y-STR Database, Release 1.0.

SWGDAM is currently considering recommending the use of the consolidated database for population frequency estimation in casework. In the reporting of matches, haplotype searches of the population database should be conducted using all loci for which results were obtained from the evidentiary sample. In cases where less information is obtained from the known sample, only those loci for which results were obtained from both the known and evidentiary sample should be used in the population database search.