D41   STR Database Technology Improvements— Storage, Retrieval, Matching, and Reporting

Scott Strong*, Jon Davoren, MS, and Edwin F. Huffine, MS, International Commission on Missing Persons, Alipasina 45a, Sarajevo, Bosnia-Herzegovina

Upon completion of this course, participants will be given an overview of the software technology demonstrating the ability to ware- house, manipulate, search, and report STR data using "Off the Shelf" Personal Computers.

STR Technology does not have to be expensive.

It is estimated that up to 40,000 individuals are still missing from the armed conflict in the former Yugoslavia during the 1990s. One of the primary missions of the International Commission on Missing Persons' (ICMP) is to help provide closure to families by aiding in identifying their missing loved ones. Various groups within the Forensic Sciences Program are producing tremendous amounts of data and a need for some specific technology to manage this data is crucial. This is especially true within ICMP's large-scale DNA testing program that produces gigabytes of data every month.

Beginning in May of 2001, the first 16 loci STR profile from a blood sample was obtained and, over the course of the next ten months, the relatively low level of data produced from minimal throughput of samples was easily stored and managed using "Off the Shelf" spread sheet software. As methodologies and the scientific expertise improved, throughput increased. These improvements aided in transforming the ICMP's DNA laboratories into a high throughput DNA processing system in which the number of bone and blood samples being tested on a monthly basis increased from 100 and 200 to 500 and 3,500, respec- tively, from February to March of 2002. This increase in data quantity and the resulting need for the management of such data meant that spreadsheet technology would no longer be efficient for the task at hand. The technology created at ICMP allows STR data to be stored in either central or local locations and allows access to the data only by those with appropriate logon credentials. This technology has been dubbed "ICMP-IWH." (International Commission on Missing Persons Information Ware House). In order to access data from ICMP-IWH software, the user must first successfully log into the system. While the ICMP currently uses a 16-plex in obtaining the majority of its STR data, the ICMP-IWH system is adaptable to meet varying needs. The potential for future increases in the number of loci being used has been dealt with by allowing up to 24 loci per sample to be entered into the system, along with the corresponding allowable values for those loci. The order in which this information is displayed, edited, and reported is also configurable. The ICMP-IWH has two ways of entering data into the database: either by manual entry or computer downloads. During manual entry, if a loci value is "Out of Range" the operator will be notified and can choose to accept or reject that particular sample. Importing STR data needs to meet certain criteria. The file must be comma or tab delimited with sample ID and STR data. In order to insure that the correct loci value is imported to its corresponding location in the database, the import file must have a "header record, " i.e., first row of data, sample ID, D3S1358, THO1, D21S11, etc. The import engine does not require the import file to have any particular order for the columns of data, but they must be consistent in arrangement for that one file. Multiple sample categories are supported, i.e., staff, blood, bone, etc. The STR matching component of this software allows the operator to choose what sample categories to compare against each other, or an indi- vidual sample to a group. There are settings to adjust the number of minimum required loci or maximum excluded loci in either a half-band or full-band sharing mode. The software will provide the operator with the included and excluded loci "Hits" on any matching report. Statistical calculations may be performed on all data or individual groups using allele frequencies from different populations as necessary. This software ICMP-IWH) runs the windows 32 bit operating systems. It has been tested on Windows95, Windows98, Windows98SE, Windows NT4.0 (sp5, sp6a), Windows2000 (sp1), and Windows XP and well as on Dell, AST, Acer and component-assembled computers. Minimum configura- tions that have been tested are PII-233Mhz with 32Mb ram and 4Gb hard drive. This software has also been tested on an Acer PIII 1Ghz (X2) SMP with 1Gb ram and 30Gb SCSI hard drives.

The development of a centralized computer system that receives all DNA data, a corresponding DNA matching program, a LIMS for the DNA system, as well as a universal bar coding system for all four of ICMP's DNA laboratories has been successfully implemented. These combined and integrated computer capabilities are an essential component to ICMP's DNA laboratory system and permit the ICMP to properly manage the gigabits of data that are produced each week and are critical when dealing with a system that produces an average of several hundred DNA match reports per month.

DNA, STR, Database