

FINDING THE NEEDLE IN THE HAYSTACK: HOW MANY LOCI DOES IT TAKE TO FIND A SINGLE INDIVIDUAL IN A DNA DATABASE OF 2 MILLION INDIVIDUALS?

Kevin C. McElfresh and Young K. Kim
The Bode Technology Group Inc., Springfield, VA



There are currently 1.8 million individuals under some form of detention or justice related supervision in the United States. Even though the 13 CODIS STR loci are the basis for felon offender DNA typing and data banking in the United States, it is clear from basic population genetics parameters that in a population of approximately 2 million, 13 loci are not needed to identify a single individual. To determine the most efficient forensic database search strategy using an individual's STR genotypes in a hypothetical United States of America felon database, bootstrapped populations were created.

The basis of the bootstrapped populations were from the state of Virginia felon/offender DNA database (as of November 1999), the North Carolina felon/offender DNA database (as of December 1999), both constructed at The Bode Technology Group, Inc. (TBTG), and the TBTG database. The samples were analyzed for genotypes of thirteen polymorphic STR loci (CSF1PO, TPOX, TH01, vWA, D16S539, D7S820, D13S317, D5S818, D18S51, D21S11, D3S1359, D8S1179, and FGA), consistent with the FBI's CODIS database requirement. Using the most frequent genotypes and observed heterozygosities in the bootstrapped database, the number of matches were determined for each multiplex system and then for the combined thirteen loci. Additional statistics generated from the databases included expected heterozygosity and the power of exclusion.

In addition, an arbitrary profile was searched one locus-at-a-time to determine the minimum number of loci necessary to find a match. As the number of loci increased from one to eight (by ascending and then descending heterozygosities), the minimum number of loci required to isolate an individual in the database using only eight loci was eight (by ascending heterozygosity) and seven (by descending heterozygosity). The minimum number of loci required to isolate an individual in the database using the thirteen loci required by CODIS was eight (by ascending heterozygosity) and seven (by descending heterozygosity). These results were based on the worst case scenarios in that the more common alleles were used to test the search strategies.