

## **Simulation Studies of Relatedness Searching in Forensic DNA Databases**

Susan Berdine

Denver Police Department, Crime Laboratory

Identification of familial relationships in CODIS databases is receiving considerable attention in the United States. This attention is based on successes in the United Kingdom and the United States in providing investigative leads, especially in violent crimes. The identification of close relatives through kinship analyses or paternity testing with statistics using the likelihood ratio approach has strong theoretical and statistical foundations.

The work described here formalizes the approach to systematically searching and detecting close relatives in DNA databases. What is needed is an intelligent way to rank these results and to limit the number of potential relative 'leads' that an investigator must examine. To achieve this, and to determine performance parameters for these types of searches, we conducted simulation studies of 500 sibling-pairs and parent-child pairs on a background of 1,000,000 randomly generated 13 locus STR-DNA profiles taken to be representative of a large forensic database. A familial search was then conducted using the software package developed and the results were ranked and plotted as the position in the list versus the number of simulated sibling-pairs detected by the search. Each simulated sibling-pair was tagged in the software to be identifiable.

Our results demonstrate that the likelihood ratio ranking of potential related pairs of individuals based on the 13 STR loci can efficiently identify relatives in forensic DNA databases. Nevertheless, ranking and sorting of the search results remains a challenge. Differences in ranked lists based on the source of allele frequency data were tested using global, U.S. and local Denver data. Both sibling and parent-child simulations were detected optimally in the first 100 positions in the ranked results list. Sibling detection was slightly better than parent-child detection, but both detect about 87% to 90% of the relatives in the first 100 positions in the ranked lists. Our analysis using positive predictive values and receiver operator curves, predicts that a likelihood ratio of approximately 50,000,000 for sibling-pairs and 5,000,000 for parent-child could be used as a general threshold for applying more resources, such as Y-STR analysis or investigative time to determine true relatedness. This research is of interest to the forensic community because as more jurisdictions have success, crime laboratories will be asked about the technology and how to implement these types of searches.