E30   Discovering Related Individuals in STR DNA Databases

Jason R. Gilder, PhD*, Forensic Bioinformatics, 2850 Presidential Drive, Suite 160, Fairborn, OH 45324; and Dan E. Krane, PhD*, Wright State University, Department: Biological Sciences, Biological Sciences Building 128, 3640 Colonel Glenn Highway, Dayton, OH 45435

After attending this presentation, attendees will have a better understanding of the makeup of allele frequency databases, issues with the presence of population substructure, how population substructure is currently being identified, and new methods of identifying population substructure that could be used for better quality control of allele frequency databases.

This presentation will impact the forensic science community by providing awareness of a new method of identifying the presence population substructure in forensic DNA databases that is far more reliable than the Hardy-Weinberg and linkage equilibrium tests that are currently being employed.

Allele frequency databases form the basis of the statistical weighting of forensic DNA profiles. An ideal allele frequency database would be representative of an underlying population and consist of randomly-chosen, unrelated individuals. Hardy-Weinberg and linkage equilibrium tests have been heavily relied upon to establish that the populations that have been sampled to establish allele frequencies are free from significant amounts of population substructure. Population substructure can lead to increased homozygosity and can misrepresent the allele frequencies present in the actual population. However, Hardy- Weinberg and linkage equilibrium are known to be weak statistical tests and are subject to the fallacy of denying the antecedent. For example, the presence of rain indicates the presence of clouds. However, a lack of rain does not indicate that no clouds are present. Loci that deviate from Hardy-Weinberg equilibrium expectations may be an indication of population substructure. However, loci that are consistent with Hardy- Weinberg equilibrium expectations may also be from populations with substantial amounts of substructure, but are not identified as such.

Kinship analyses in which each individual in a database is compared with every other individual to identify pairs of possible relatives is a more sensitive means of assessing population substructure. While Hardy-Weinberg and linkage equilibrium are intended to identify overall population substructure, kinship approaches have the potential to identify smaller clusters of related individuals that might skew allele frequencies, yet are unlikely to be identified by tests for Hardy-Weinberg or linkage equilibrium.

For this study, the Federal Bureau of Investigation (FBI) and National Institute of Standards and Technology (NIST) African American, Caucasian, and Southwest Hispanic allele frequency databases were examined with kinship analyses to determine if they contain pairs of individuals that are likely to be close relatives. Both the FBI and NIST population databases appear to contain pairs of closely related individuals. An exploratory study of Hardy-Weinberg and linkage equilibrium was also carried out to characterize the limits of the ability of these tests to identify the presence of related individuals in a database.

DNA Database, Population Substructure, Kinship Analysis