Statistical inferences on Genetic Data from Challenged Samples: From Low Copy Number Human Identification to Attribution in Microbial Forensics

Bruce Budowle, Department of Forensic and Investigative Genetics, Institute of Investigative Genetics, University of North Texas health Science Center, Ft Worth, Texas

Genetic typing is a powerful means to establish identity in criminal cases where biological evidence is found at crime scenes, for paternity testing and inheritance matters, identification of victims in mass disasters, and identification of missing persons from human remains, as well as providing investigative leads in cases of bioterrorism and biocrime. Conveying the significance of a failure to exclude is an important part of the interpretation of the results of an analysis. The statistical approaches for human single source, mixture, and kinship analyses are based on principles of population genetics and sampling statistics and include issues such as Hardy-Weinberg and linkage equilibrium expectations, correction for effects of population substructure, minimum allele frequencies, lineage-based counting methods, and population data sets. While the statistical approaches for inferences of the weight of evidence are well-established, new situations have arisen that require additional thought regarding placing significance on a result of failure to exclude. Two areas are Low Copy Number (LCN) typing and Microbial Forensics and the statistical issues regarding their applications will be raised and discussed.

LCN typing refers to the analysis of any sample that contains less than 200 pg of template DNA (~33 human diploid cells). Exaggerated stochastic sampling effects are observed when amplifying a small number of starting templates during the PCR. The result is that several phenomena can occur: a substantial imbalance of two alleles at a given heterozygous locus, allelic dropout, and/or increased stutter. With increased sensitivity of detection there also is a concomitant increased risk of contamination. The approach most widely used for the designation of an allele in a LCN sample requires the division of the sample into two or more aliquots and reporting only the alleles that are common in at least two replicates. However, the suppositions and degree of confidence surrounding such practices are not well-defined. Moreover, the approaches used may have some application for single-source samples but are far more complex for mixtures. Therefore, until better developed it may be better to use LCN typing for identification of missing persons and for investigative leads.

Microbial forensics is an evolving sub-discipline of forensic science for analysis of evidence from a bioterrorism act, biocrime, hoax or an inadvertent release for attribution purposes. Some practices will be similar to that applied to human DNA analyses, such as Quality Assurance, generation of databases, and conveying significance of a nucleic acid analysis result. However, most of the approaches for developing statistical inferences will be very different for microbial forensics because of the lineage-based markers (more like Y chromosome and mitochondrial DNA markers) and additional genetic inheritance characteristics, such horizontal gene transfer. The databases generated will be very different.

Moreover, the power of discrimination (or ability to individualize) will almost always not be able to reach that enjoyed for human identity testing, instead a most recent common ancestor approach will be considered. Additional major limitations will be the unknown history of samples and host-pathogen interactions.