THE VALIDATION OF A STATISTICAL TOOL FOR THE ANALYSIS OF DNA MIXTURES

Rebecca Mikulasovich, MS, F-ABC, Michael Donley, MS, F-ABC, Dustin Foley, MS, F-ABC, Katelyn Weymouth, PhD, Mark Powell, MS, F-ABC, Katie Welch, MS, F-ABC, Roger Kahn, PhD, F-ABC, Harris County Institute of Forensic Sciences Genetic Laboratory

Gone are the days when DNA analysis was only brought to bear on high yield blood and semen samples from homicides and sexual assaults. Improvements in time of analysis, sensitivity and cost mean it is not uncommon to encounter touch DNA samples in the crime lab. The DNA profiles obtained from touch samples are often complex mixtures that are not interpretable using traditional mixture interpretation guidelines and statistics. Laboratories that invest resources to process these samples must utilize alternate methods in order to take full advantage of the data obtained.

The analysis of complex mixtures comprised of partial DNA profiles where alleles may be missing (allelic dropout) or where additional alleles may be present (drop-in) has proven to be a particularly difficult problem to solve. At the Harris County Institute of Forensic Sciences, the R based statistical software package developed by Hinda Haned called Forensim was validated for these complex mixtures. Several models have been proposed in the literature to overcome complex mixture interpretation issues but most are generally not available in an open source platform. The specific module of interest in Forensim is LRmix, which follows the method of Gill et al. published in Forensic Science International Genetics 2007;166(2–3):128–138. It enables the calculation of likelihood ratios for complex STR profiles with allele drop-in, dropout and multiple contributors.

Initial validation work focused on the calculation of the probability of drop-out and drop-in. These two probabilities are entered manually before running the LRmix module. Using validation samples of varying concentration, several published methods including the counting method, tailed method and logistic regression were applied and an in-house probability of drop-out of 0.14 was determined. To determine the probability of drop-in, hundreds of casework negative controls were analyzed at instrument baseline looking for spurious allele peaks. Only one peak was detected leading to a probability of drop-in of < 0.01. Because 0.01 is the smallest probability of drop-in allowed by LRmix, this is the value utilized. Another tool available in the LRmix module is a Monte Carlo simulation of probability of drop-out that is done after the profile is evaluated and an assumption of the number of contributors is made. This simulation was run throughout the validation and in all scenarios the calculated probability of drop-out was within the simulated range. This confirms the calculated probability as reasonable and these simulations will also be included in casework.

Forty two person and three person complex mixtures were evaluated using LRmix to demonstrate sensitivity, reproducibility and precision. All true contributors to the mixtures were correctly associated when evaluated against the mixtures. Over 100 manually random generated single source profiles were created and all were correctly excluded when evaluated against the mixtures. LRmix is also capable of automating this step using the performance check module where random profiles are created based on the population database used and compared to the mixture profile. The user chooses how many profiles to compare. During the validation 100 to 10000 random generations were compared and for casework purposes 1000 will be used. This check will provide an idea of how likely it is a random person would be included or excluded from the mixture.

Overall, the validation of LRmix demonstrates it is an acceptable method for the calculation of the likelihood ratios for complex two and three person mixtures and is an economical tool that can be incorporated by any laboratory.