B136 Evaluating the Forensim Package for Modeling DNA Mixtures for Statistical Calculations of Contributor Attribution in North American Populations

Nicole M.M. Novroski, PhD*, University of Toronto, Mississauga, ON L5L1C6, CANADA; Ashley Moo-Choy, University of Toronto Mississauga, Mississauga, ON L5L 1C6, CANADA; Frank R. Wendt, PhD*, Yale School of Medicine, West Haven, CA 06516

Learning Overview: After attending this presentation, attendees will understand the functionality and efficacy of the forensim R package for application to forensically relevant DNA mixture simulations.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by providing an external evaluation of the functionality of the forensim R package to generate DNA mixture profiles and genotypes based on population-based Short Tandem Repeat (STR) allele dynamics. As laboratories improve DNA mixture deconvolution efforts, there is a critical need to evaluate kit deconvolution capabilities using modeling programs and algorithms in addition to benchtop artificial mixtures. These approaches leverage population dynamics and ensure that this study's conclusions are accurate and reproducible in casework settings.

Statistical modeling and simulations are common tools employed in forensic genetics to assess the validity and accuracy of DNA mixture deconvolution approaches. There is growing interest in understanding how individual genotypes and alleles from different population groups can be over or underrepresented in DNA mixture profiles. Using mixed ancestry mixtures, it has been demonstrated that populations with lower genetic diversity are overrepresented in mixture deconvolution match statistics. There remains some uncertainty however, as to how minor contributors match to low diversity populations in models assuming admixed major and minor contributors. Furthermore, systematic ancestry proportions remain underinvestigated; this study hypothesizes that systematically increasing the African versus European ancestry proportion of minor contributors (i.e., 5%–50% in 5% increments) will reveal biases in minor contributor match statistics influenced by African ancestry proportion. As complex DNA mixtures are more routinely interpreted, it is important to understand how statistical modeling and simulation(s) represent genotype frequency distributions in the ancestral affiliation of each mixture contributor.

In this study, a detailed evaluation was conducted to determine whether the forensim R package is capable of modeling complex DNA mixtures from known population data. In R, forensim provides a series of simulation tools to generate genetic mixtures such as those encountered in forensic casework. Genotypes are derived from population allele frequencies and population structure. Further, the package also allows for the interpretation of complex Short Tandem Repeat (STR) results by providing a maximum likelihood estimator for the number of contributors to a mixture, random man non-excluded probabilities, and likelihood ratios for low template samples with drop-out, drop-in, and varying number of contributors and replicates. Recent studies using forensim demonstrate that underrepresented population(s) (i.e., those populations with lower genetic diversity) may be overrepresented in DNA mixture match statistics due to limited allele diversity in their population datasets. Of particular interest, DNA mixture likelihood ratios were demonstrated to suggest that non-contributors from lower genetic diversity populations are included in the mixture(s), so a detailed investigation into the possibility of false positive(s) and negative(s) in a series of simulated DNA mixtures from five populations has been performed. Using known genotypes and allele frequencies from previously published population datasets, this study compared genotypes, heterozygosities, and allele combinations to evaluate model accuracy given variable input population information. Likelihood ratios were calculated to assess whether non-contributors were included within the DNA mixture profiles as a measure of quantifying false positive rates across population groups and number of contributors.

DNA Mixtures, Mixture Simulations, Allele Frequencies