E4 Empirical Analysis of the STR Profiles Resulting from Conceptual Mixtures

Dan E. Krane, PhD, Wright State University, 3640 Colonel Glenn Highway, Biology, Dayton, OH 45435*

After attending this presentation, attendees gain proficiency towards reliably determining the number of contributors within a mixed DNA sample based on the maximum number of alleles observed at any given genotyped locus.

This presentation will impact the forensic community by demonstrating that analysts often underestimate the number of contributors within a mixed DNA sample and by illustrating steps for correcting that inaccuracy.

Samples containing DNA from two or more individuals can be difficult to interpret. Even ascertaining the number of contributors can be challenging and associated uncertainties can have dramatic effects on the interpretation of testing results. Using an FBI genotypes dataset, containing complete genotype information from the 13 Combined DNA Index System (CODIS) loci for 959 individuals, all possible mixtures (146,536,159) of three individuals were exhaustively and empirically computed. Allele sharing between pairs of individuals in the original dataset, a randomized dataset and datasets of generated cousins and siblings was evaluated as were the number of loci that were necessary to reliably deduce the number of contributors present in simulated mixtures of four or less contributors. The relatively small number of alleles detectable at most CODIS loci and the fact that some alleles are likely to be shared between individuals within a population can make the maximum number of different alleles observed at any tested loci an unreliable indicator of the maximum number of contributors to a mixed DNA sample. Peak heights and areas sometimes provide additional data that is useful for the purpose of mixture resolution but this information is not utilized in the analysis presented here.

Instead, this study examines the interpretation of STR data in cases where this information is unreliable (i.e. when degradation has occurred and/or stutter complicates interpretation), unavailable (i.e., only a laboratory's summary report is provided for review) or uninformative (i.e., the relative contributions by two or more contributors are similar). As a result, the study represents a worst-case analysis of mixture characterization. Within this dataset, approximately 3% of three-person mixtures would be mischaracterized as two-person mixtures and more than 70% of four-person mixtures would be mischaracterized as two- or three-person mixtures using only the maximum number of alleles observed at any tested locus.

We also assess the ramifications of invoking analyst discretion to discard a seemingly anomalous locus by determining the number of three-person genotype mixtures where discarding a single locus with the highest number of different observed alleles produces results consistent with mischaracterization of the mixture as a single source sample or as a two-person mixture. A larger amount of pair-wise allele sharing was observed between individuals in the original FBI dataset relative to five datasets of randomized individuals suggesting that the FBI dataset may contain some pairs of closely related individuals. The extent of allele sharing between siblings in large-scale simulations also suggests that perfect 13 locus matches (26 out of 26 possible alleles) occur at a frequency (an average of 3.0 per 459,361). This frequency suggests that some are likely to exist in large populations such as the general population of the United States and even eventually in DNA profile datasets that contain large numbers of close relatives.

DNA, Mixture, Bioinformatics

 ** Presenting Author*