B37 Alleles With Dissimilar Frequencies Between Ethnic Populations Increase the Uncertainty in DNA Mixture Interpretation

Sam Kwiatkowski, PhD, Harris County Institute of Forensic Sciences, Houston, TX 77030; Michael A. Donley, MS, Houston, TX 77021; Katherine Welch, MS, Harris County Institute of Forensic Sciences, Houston, TX 77054; Roger Kahn, PhD, Harris County Institute of Forensic Sciences, Houston, TX 77054-2001*

Learning Overview: The goal of this presentation is to communicate this study's results after evaluating the impact of assigning different allele frequency databases to contributors of two-person mixtures.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by demonstrating that Likelihood Ratios (LRs) produced from the current practice may be non-conservative when mixtures comprise alleles with substantially dissimilar frequencies between populations.

The current practice in the statistical analysis of mixed DNA profiles is to compute LRs assuming the same population database for each contributor. This practice does not consider that mixtures may comprise individuals best represented from different populations. LRs reflect more than the comparisons to a Person Of Interest (POI); they also consider the genotypes proposed for any unknown contributors. When the genotypes proposed for contributors favor different populations, it is possible for outcomes to be non-conservative. Although past studies have suggested that the impact of this uncertainty is minor, the impact of alleles with large differences in frequencies between relevant populations has not been considered. For example, the Harris County Institute of Forensic Sciences calculates LRs using three Federal Bureau of Investigation (FBI) populations (African American/Bahamian/Jamaican, Caucasian, and Southwest Hispanic populations), which include alleles showing substantial differences in frequencies; some alleles have frequencies ranging 1%–7.5% in one population while having zero observations in another. It was expected that LRs would be increased when rare alleles are present in mixtures.

Mixtures were deconvoluted using the probabilistic genotyping system STRmix™. Random non-contributor profiles producing LRs greater than 100 were identified using propositions: (H_1) the DNA originated from the POI of a first population and an unknown contributor of a second population; and (H_2) the DNA originated from an unknown contributor of the first population and an unknown contributor of the second population. Same- and different-population LRs were calculated based on the methods of STRmix™. The methods were adapted for calculating different-population LRs by considering zero coancestry between alleles sampled under different populations. The populations considered were the FBI African American/Bahamian/Jamaican, Caucasian, and Southwest Hispanic populations. Nine sub-source LRs were produced for each non-contributor profile. The most conservative same-population LR was compared to the most conservative LR over all nine outcomes. The impact of specific alleles having substantially dissimilar frequencies between populations was evaluated. Candidate alleles were those having at least about 100-times greater posterior mean frequency in one population over another.

Comparisons of 7,980,000 were made between two-person mixtures and random non-contributor profiles. Mixtures lacking candidate alleles showed a wide variation in outcomes resulting from considering different populations: 7.1% of non-contributor LRs were reduced by at least 90% in magnitude (largest observed was 97.0%), 71.2% were reduced by 10%–90%, and 21.7% were reduced by less than 10% or unchanged. The presence of candidate alleles decreased the proportion of LRs that were reduced by 10%–90%: the proportions for mixtures having one or two candidate alleles were 26.8% and 5.4%, respectively. This was accompanied by increases in the proportions of LRs that were reduced by at least 90% (58.5% and 21.6% for mixtures comprising one or two candidate alleles, respectively) or less than 10% (14.6% and 73.0%, respectively). Several LRs were reduced by at least 99%; the proportions for mixtures comprising one or two candidate alleles were 9.8% and 16.2%, respectively.

When mixtures comprised a single candidate allele, the proportion of LRs reduced by one or more orders of magnitude was increased. When mixtures comprised multiple candidate alleles, a proportion of LRs were reduced by multiple orders of magnitude. Candidate alleles having the greatest impact were those with frequencies ranging about 1%–7% in one of the FBI African American/Bahamian/Jamaican, Caucasian, and Southwest Hispanic populations and zero observations in another population. Due to the large expected frequencies of candidate alleles and their impact on LRs, it may be necessary to report statistics for considering whether mixtures comprise contributors of different populations.

Probabilistic Genotyping, STRmix™, Mixture Interpretation

---

*Presenting Author