B139 Validating TrueAllele® Interpretation of DNA Mixtures Containing up to Ten Unknown Contributors

Nasir A. Butt, PhD*, Cuyahoga County MEO, 11001 Cedar Avenue, Cleveland, OH 44106; David W. Bauer, PhD*, Cybergenetics, Pittsburgh, PA 15213; and Mark W. Perlin, PhD, MD, Cybergenetics, 160 N Craig Street, Ste 210, Pittsburgh, PA 15213

After attending this presentation, attendees will understand the reliable interpretation of low-level DNA mixtures containing many people. Fully Bayesian computer methods can accurately extract identification information from complex DNA evidence.

This presentation will impact the forensic science community by providing empirical support for sophisticated scientific interpretation of DNA evidence often found in criminal justice applications.

Mixed samples of known genotype composition were prepared following a randomized experimental design that included very low contributor amounts. The mixture samples contained from two to ten contributors. The Cuyahoga County Regional Forensic Science Laboratory (Cuyahoga) amplified the samples using a PowerPlex® Fusion Short Tandem Repeat (STR) kit. Cybergenetics and Cuyahoga independently conducted TrueAllele® casework testing to interpret the DNA mixture data.

TrueAllele® provides fully Bayesian analysis. The system examines all STR data without human intervention, without knowing the comparison genotype; data objectivity eliminates contextual bias. The computer extracts all information from the evidence data, inferring genotypes and other parameters; the model completely eliminates unneeded calibration. Inferred evidence genotypes are compared with references only after genotypes have been separated from DNA mixture data; the two-phase Likelihood Ratio (LR) approach enforces objectivity.

The study examined sensitivity, specificity, and reproducibility of TrueAllele® match information. Fewer contributors in a mixture generally improved these metrics. With more contributors, longer Markov Chain Monte Carlo (MCMC) statistical sampling increased accuracy. Assuming excess contributors did not affect match information for true contributors, but added superfluous genotypes containing little information.

Sensitivity: As contributor number increased from two to six, average match strength decreased from 24 to 5 LR log units ("ban"). With up to four contributors, only positive log (LR) values were observed.

Specificity: As contributor number increased from two to six, average specificity decreased from -33 to -13 ban. When comparing with non-contributors, a match statistic of more than a thousand had a false positive probability of under 0.0001. With LR more than a hundred thousand, the probability decreased to less than 0.000001.

Reproducibility: Fewer contributors conferred greater reproducibility. The within-group standard deviation for two contributors was 0.17 ban, and for six contributors was 1.00 ban.

DNA Amount: Match information depended on a contributor's DNA amount, regardless of how many contributors were in the mixture. (More contributors mean less DNA per contributor.) Contributors comprising less than 10% of the mixture exhibited an average log (LR) of 3 ban; however, with fewer contributors, the match statistics were higher. Major contributors (more than 50%) averaged 28 ban. Contributors over 20% gave positive log (LR) values, regardless of how many contributors were present.

Contributor Number: When given the number of contributors a user observed in the data, the computer's mixture solutions were better than when given the expected number known from the study design. This result demonstrated application robustness, as practiced by forensic analysts on DNA mixtures.

Assumed Genotypes: With low-level minor contributors, providing the computer with known (i.e., previously matched) references reduced uncertainty in genotype inference. For a 15% minor in a ten-person mixture, the log (LR) value increased from 3 to 8 ban when assuming previously matched references as known genotypes. The log (LR) value for a 2% minor in a ten-person mixture increased from 1 to 4 ban. When assuming known genotypes, specificity improved — the non-contributor log (LR) average shifted from -3 to -6 ban.

MCMC Sampling: More MCMC sampling increased sensitivity. With six contributors, MCMC sampling at five thousand cycles gave an average log (LR) of 2.4 ban for true contributors; this increased at one hundred thousand cycles to 5.2 ban. Specificity was essentially unchanged by further MCMC sampling, so faster run times did not increase false positives. Regardless of contributor number, reproducibility at one hundred thousand cycles revealed a between-run variability of less than 1 ban.

Independent Testing: Cybergenetics average match statistic for true contributors was 8.52 ban, while Cuyahoga's average was 8.24 ban. On average, the log (LR) values for independent comparisons conducted at the two sites were within 1.15 ban.

A validation study showing reliable interpretation of complex DNA evidence will be presented.

DNA Evidence, Mixture Validation, Bayesian Computing

*Presenting Author