



A70 Investigative DNA Databases That Preserve Identification Information

Mark W. Perlin, PhD, MD*, *Cybergenetics, 160 North Craig Street, Suite 210, Pittsburgh, PA 15213*

After attending this presentation, attendees will understand how investigative DNA databases can be improved to better preserve DNA identification information. When databases store and match probabilistic genotypes, instead of allele lists, all the biological evidence information can be preserved.

The presentation will impact the forensic science community by enabling DNA databases to make better use of biological evidence for investigations. By moving to a more informative probabilistic genotype representation, databases like CODIS can become far more sensitive and specific. This sharpened information capability makes DNA databases more successful in connecting the right criminal to DNA mixture evidence.

A DNA database can link crime scenes to suspects, providing investigative leads. These DNA associations can solve cold cases, track terrorists, and stop criminals before they inflict further harm. However, current government databases do not fully preserve DNA identification information and cannot maximize public safety.

DNA data is summarized in a genotype. The genotype can be stored on a database, and compared with other genotypes to form a likelihood ratio (LR) match statistic. Data uncertainty, present in most evidence, particularly DNA mixtures, translates into genotype probability.

Highly informative DNA mixture interpretation uses all the quantitative data, placing higher probability on more likely genotype values. Most evidence, though, is currently interpreted by a qualitative human review that diffuses probability across infeasible solutions. Since the LR is proportional to the true genotype probability, weaker interpretation methods lead to weaker (or nonexistent) DNA matches.

The weakest DNA mixture interpretation method is the Combined Probability of Inclusion (CPI), also known as Random Man Not Excluded (RMNE). CPI uses thresholds to truncate quantitative data into all-or-none qualitative "allele" events. The current DNA databases (including CODIS) use a CPI allele representation that discards considerable genotype information, losing sensitivity and specificity.

The "probabilistic genotype" representation described by SWGDAM¹ is part of the new ANSI/NIST-ITL data exchange standard.² Unlike allele lists, a probability representation can preserve all DNA evidence identification information on a forensic database and calculates accurate LR statistics as it matches across the database.

ISFG's 2006 mixture guidelines³ recommend the more informative LR over CPI. Unfortunately, current databases transform hard won LR genotypes into less informative CPI alleles. This presentation will explain how genotype probability can preserve identification information for DNA investigation.

Forensic DNA is an information science with DNA databases having the potential for considerable identification power. However, current database implementations discard most of the information in DNA mixture evidence. This presentation will help practitioners understand how to build and use investigative DNA databases that preserve all of the identification information present in their biological evidence.

References:

- ¹ Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, Morling N, Prinz M, Schneider PM, Weir BS. DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci Int.* 2006;160:90-101.
- ² SWGDAM. Interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories. 2010; Paragraph 3.2.2 (probabilistic genotypes).
- ³ Carey S. Data format for the interchange of fingerprint, facial & other biometric information, ANSI/NIST-ITL 1-2011. Gaithersburg, MD: American National Standards Institute (ANSI) and National Institute for Standards and Technology (NIST) 2011; Sections 18.020-18.021 (probabilistic genotypes).

DNA Database, DNA Mixture, Likelihood Ratio