



B185 Genotype Information Criteria for Forensic DNA Databases

John Donahue, MA, Beaufort County Sheriff's Office, Beaufort, SC 29906; Mark W. Perlin, PhD, MD, Cybergenetics, Pittsburgh, PA 15213*

Learning Overview: The goal of this presentation is to help the attendee better understand effective probabilistic genotyping for DNA database searches.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by demonstrating that an automated probabilistic method for searching DNA databases can use more DNA profile information to increase the number of solved cases with minimal human review effort.

In Bayesian statistics, Kullback-Leibler (KL) divergence is a measure of data information gain from prior probability to posterior probability. Bayesian methods can be applied to forensic DNA mixtures. The KL then estimates the match information present in a contributor's probabilistic genotype. Thus, KL can be used to predict the match statistic when a reference sample is not available.

The Combined DNA Index System (CODIS) is a network of DNA databases shared by law enforcement agencies across the United States. It is primarily used for comparing DNA profiles—crime scene evidence with convicted offenders or arrestees. The CODIS software searches for matches by directly comparing the alleles of two profiles. Depending upon the search criteria, a one-allele profile difference may cause a mismatch that precludes match reporting. In practice, few mismatches occur when comparing highly certain genotypes, such as those derived from a single source evidence sample or an obvious major contributor to a DNA mixture.

However, direct allelic comparison can create challenges when evaluating more complex DNA mixtures. These typically have more genotype uncertainty than single source samples. A direct comparison search of a DNA mixture having several alleles at multiple loci can produce many adventitious candidate matches.

To reduce such false positive results, the National DNA Index System (NDIS) Operational Procedures require that forensic mixtures submitted to NDIS contain at least eight of the original thirteen CODIS core loci, with a maximum of four alleles at any locus. In addition, the mixture must have a match rarity estimate no greater than one in ten million, calculated at moderate stringency for the original CODIS core loci. These restrictions do help reduce adventitious matches. But they also prevent the upload of mixtures that fail eligibility criteria, even if they contain considerable match information. It is therefore likely that current NDIS eligibility criteria prevent the identification of true matches to complex DNA mixtures, which would leave some crimes unsolved.

This presentation describes the database search methods employed by a local laboratory that has a TrueAllele® computer system for developing probabilistic genotypes from complex DNA mixtures. By calculating the KL of a probabilistic genotype, the TrueAllele® computer measures potential match information across all the data, not just at a selected locus subset. The lab uses KL match prediction to assess inferred genotypes for database search suitability and help determine alleles for upload. After a database search returns potential matches, the computer automatically calculates match statistics between the retrieved references and the original inferred mixture evidence.

Direct CODIS allele comparison can return multiple reference genotypes. However, the laboratory's automated TrueAllele® comparison reduces human review time to a few minutes. Moreover, the process provides objective and unbiased match evaluation. This advanced DNA database computing approach has enabled the upload of previously unsearchable profiles to the State DNA Index System (SDIS) and has solved previously unsolvable criminal cases. The results suggest that NDIS might achieve increased database success through KL-directed probabilistic genotyping search.

CODIS, Probabilistic Genotyping, Match Estimation