Abstract 18

INTERPRETING UNCERTAIN DNA EVIDENCE: A NEW STATISTICAL APPROACH

Mark W. Perlin, Ph.D., M.D., Ph.D.[1], Joseph B. Kadane, Ph.D.[2] and Robin W. Cotton, Ph.D.[3]
[1]Cybergenetics, Pittsburgh, PA
[2]Statistics Department, Carnegie Mellon University, Pittsburgh, PA
[3]Biomedical Forensic Sciences, Boston University, Boston, MA

DNA typing is a powerful method of comparing biological specimens and establishing genetic identity. However, in actual forensic casework, there is typically considerable uncertainty in the underlying quantitative DNA data. This data uncertainty is currently addressed by qualitative "include/exclude" interpretation methods which discard much of the identity information present in the data. These methods can be improved upon through the use of a quantitative likelihood function which models the data. This paper introduces a general probability framework for describing DNA interpretation methods, and compares the identification power of leading methods on representative uncertain evidence. We demonstrate that better modeling of the data can greatly improve DNA identification.

The previous standard of forensic human identification was the fingerprint. As with DNA, it is well established that no two individuals (even identical twins) share the same fingerprint. Hence, comparing pristine reference fingerprints is a virtually infallible way to establish human identity. However, crime scene prints do not provide such perfect data, since the images can be partial, blurred, distorted or otherwise uninformative. Fingerprint experts have recently made incorrect matches involving uncertain data which have demonstrated the limitations of the method in casework applications.

As with fingerprint technology, DNA identification is based on the fact that there is considerable variation in the profiles of different individuals. And, like fingerprints, actual DNA data arising in casework applications can have considerable uncertainty. DNA evidence can be partial, faint, shifted, contaminated, or derived from more than one individual.

In this last situation, a DNA mixture can bear little correspondence to the pristine data people usually associate with DNA evidence. Instead, the standard interpretation and match methods become unworkable, and entirely different approximations must be used, producing weaker (or incorrect) conclusions about genetic identity. Indeed, different DNA analysts often infer entirely different results from the same DNA mixture data. The subjective variation is so great that DNA labs do not even bother to validate their mixture interpretation methods in order to establish reliability.

This paper addresses the breakdown of classical DNA interpretation theory in the presence of uncertain data. It is difficult to extend the simple theory for idealized data to the complex casework situation without losing some identification information. The purpose of this paper is to provide a new theoretical framework for objectively interpreting and matching uncertain DNA evidence. We present a unified method which

can preserve the DNA match information present in the data, and which is amenable to computer implementation.