A162 Mixed DNA Analysis

Matthew J. Bohn, PhD, Analytic Services, Incorporated, 1330 Inverness Drive, Colorado Springs, CO 80910; and Michael J. Salyards, PhD, 45 High Street, Sharpsburg, GA 30277*

After attending this presentation, attendees will learn how statistically significant information can be extracted from mixed DNA samples. Some statistical techniques are relatively simple to implement such as making random match probabilities on a mixed DNA sample given an individual's DNA profile. Other more difficult techniques will be interesting to the warfighter attempting to capture a terrorist cell or a narcotics unit attempting to track a drug gang. These more sophisticated techniques extract minor commonalities in the DNA profiles using the techniques of Principal Component Analysis (PCA) and clustering algorithms.

This presentation will impact the forensic science community by demonstrating mathematical techniques that can be used to extract information from DNA samples containing more than one contributor.

Currently, mixed DNA samples face difficulties linking an individual to a crime scene or location. However, in an expeditionary environment faced by a military force, evidence linking a group of individuals could be of monumental importance. As a hypothetical example, a recently confiscated cell phone contains DNA from three individuals of al-Qaida and a mixture of four random, innocent individuals can be found to be statistically similar to a mixed DNA sample found on an IED. Cluster analysis can be used to separate the group of al-Qaida members from the random, innocent individuals. In addition, statistics can be used to ascertain the random match probability of an individual in the group. For example, the warfighter could ascertain with what probability the individual holding the cell phone is a member of the al-Qaida group that built the IED.

Analysis techniques will be discussed for mixed DNA from crime scenes, confiscated material, and opportunistic collections. A wealth of intelligence information hidden exists within mixed DNA samples. World-wide analysis and databasing of DNA mixtures will lead to increased awareness of the world-wide threat while simultaneously providing actionable intelligence to the commander in a local environment. Statistical analysis of mixed DNA samples using Principal Component Analysis (PCA) and cluster analysis enables tracking, group identification, and probability analysis of individuals in the group. Computer simulations will be presented demonstrating the exciting potential for the use of DNA mixtures. Based on a NIST database, a computer program was developed to generate random DNA profiles representative of the allelic frequencies in the database. These profiles could then be mixed and a random match probability of a random individual to the group calculated. The random match probability is then plotted versus the number of contributors. One example from the data is that with ten contributors to a mixed profile, the probability of a random individual's profile appearing in the mixture is one out of ten. With four contributors, the random match probability drops to one in 3,333. Obviously these numbers are far from the typical one in five billion random match probabilities for a single profile; however in an expeditionary setting, they may help to uncover a terrorist cell.

Given sufficient data, groups can also be identified and tracked in mixed DNA samples. A collection of 60 computer generated samples were analyzed. The 60 samples consisted of three groups: (1) completely random; (2) four random profiles plus three consistent profiles; and, (3) four random profiles plus three different, consistent profiles. The concept being modeled is that three members of a terrorist cell or drug gang might all touch a cell phone, which is then thrown away and touched by four innocent individuals. This leads to an aggregate DNA sample that contains a mixture of seven individuals. Can these groups be identified and separated? Yes, with reasonable probability, the method of PCA coupled with cluster analysis can separate the groups. In addition, the total number of alleles in these mixtures was analyzed using a Monte Carlo technique and depending on the total number of alleles, an estimate can be made on the total number of contributors.

The opinions or assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

Mixed DNA, Principal Component Analysis, Cluster Analysis