MIXTURE STR/DNA INTERPRETATION USING ALLELE PEAK AREA INFORMATION WITH THE LEAST SQUARE METHODOLOGY

Tsewei Wang[1], Ning Xue[1], Mark Rader[2], and Doug Birdwell[2]
Laboratory of Information Technology
[1]Department of Chemical Engineering, The University of Tennessee, Knoxville, TN
[2]Department of Electrical and Computer Engineering, The University of Tennessee, Knoxville, TN

Genetic samples left at crime scenes involving rape often are mixtures containing DNA from both the victim and the perpetrator. It is of paramount importance that the genetic identity of the two contributors be resolved with a high degree of confidence. Thus, the probability of a suspect being the actual perpetrator can be evaluated; or in the absence of suspects, a database search can be conducted for matching DNA profiles. Recently, reports have emerged that use the allele peak area information to determine the most probable allele assignments to each of the two contributors of a mixture. The methodology reported to date is indirect, and involves searching for a solution that minimizes the data fitting error over a discrete set of assumed mass ratio values, as well as searching through all possible allele assignment combinations. This approach is rather laborious involving iterations.

We present a new approach to determine the most probable genotypes of the two contributors using a least-square methodology. This approach is direct and elegant, in that in one step, the most probable mass ratio coefficients for the two DNA sources is computed based on the allele peak information. For each locus, ranking of the top most probable allele assignments can be made, and consistent top ranked profile across all loci indicate the most probable genotype assignments to the two contributors of the mixture. For this problem, the least square solution methodology can be posed such that minimum familiarity with mathematic principals and manipulations on the part of the caseworkers are required. Just a few steps of simple arithmetic calculations are required to follow this algorithm to determine the most probable genotype combinations. This approach has been tested extensively using simulated allele peak data in mixtures with positive results. Limitations with this approach will also be presented. Testing with real (nonforensic) mixture DNA data is planned.

*Research is supported by the FBI, under Contract: J-FBI-98-083. The views and conclusions contained in this presentation are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the US Government.