

THE PROBABLE NUMBER OF CONTRIBUTORS TO A STR DNA MIXTURE

T. W. Wang, P. Kalet, J. Pendleton, K. Gilbert, C. Lucas, and J. D. Birdwell
Laboratory for Information Technologies, College of Engineering
The University of Tennessee, Knoxville, TN, 37996-2200
wang@lit.net

In analyzing a multiple contributor mixture STR DNA sample, the first question to ask is the number of contributors. Mixture resolution into component profiles and calculation of the associated probabilities are dependent on knowing the number of contributors. In a multi-person mixture, allele dropout may have occurred; masking of minor contributor's signal by artifact peaks or noise may also occur, thereby altering the allele count distribution pattern. One relevant question to ask is: Given the mixture sample at hand, what is the probability that it is an m-person mixture, as opposed to an n-person mixture where m and n are positive integers. For instance, in a mixture sample in which no locus contains more than 4 alleles, the question becomes, given this mixture sample, what is the likelihood that it comes from a 2-person rather than a 3-person source. A related question is: Assuming it is a 3-person mixture, what is the probability that AT LEAST 1 out of the 13 core loci will contain more than 4 alleles.

Results presented here document the study of allele count distribution patterns among loci of multiple contributor mixtures, ranging from 2 to 6 persons. Based on the allele frequency data published by Budowle for Caucasians¹, various probabilities of allele distribution patterns have been derived to help forensic DNA analysts to analyze the number of contributor problem, resolve the mixtures at some loci, and connect the mixture profile to other forensic evidence related to the same crime. The presentation includes the following results:

1. Given the mixture is an n-person mixture ($n=1, \dots, 6$), the probability that each of the 13 core loci contains 1, 2, ... or $2n$ alleles/locus.
2. For 2 to 6 person-mixture samples, the probability that at least 1 locus will contain more than $2(n-1)$ allele where n denotes the number of contributors, e.g. for a 3-person mixture, the probability that at least 1 of the 13 core loci will contain more than 4 alleles, i.e. 5 or 6 alleles.
3. Likelihood calculation of the sample being a 2-person mixture as opposed to a 3-person mixture, given a 13-loci profile in which no locus has more than 4 alleles.
4. Histogram of random match probabilities of 1 million mixture samples in which no locus has more than 4 alleles assuming the number of contributors is 2, vs that assuming the number of contributors is 3.
5. Type I and Type II error calculation in hypothesis testing of a mixture sample being a 2-person mixture against a 3-person mixture.

Results reveal several very interesting observations relevant to the analysis and interpretation of a mixture sample. For instance, on average, only 5.5% of 3-person mixtures resemble 2-person mixtures (i.e. all loci contain 4 or less alleles/locus); 77% of 4-person mixtures resemble those of 3-person mixtures; whereas 99% of 5-person mixtures can be confused with 4-person (60% of the time) or a 3-person (40% of the time); almost all 6-person mixtures resemble those of 5 person mixtures (6% of the time), or 4-person mixtures (80% of the time) or 3-person mixture (14% of the time) mixture. A likelihood study reveals that for a 13-loci profile in which no locus contains more than 4 alleles, and assuming the occurrence of a 2-person mixture occurs 90% of the time over all mixture samples, then it is one million times more likely, on average that the mixture comes from a 2-person source than from a 3-person source. Explanations for some of these observations will also be presented.

¹ Budowle, B., B. Shea, S. Niezgodna, and R. Chakraborty. CODIS STR Loci Data from 41 Sample Populations, J. Forensic Sci., 2001:46(3), 453-489.