Forensic Science International: Genetics 33 (2018) 24–32

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Short communication

Estimation of the number of contributors of theoretical mixture profiles based on allele counting: Does increasing the number of loci increase success rate of estimates?

Gina M. Dembinski[a], Carl Sobieralski[b], Christine J. Picard[a,c,*]

[a] Department of Biology, Indiana University-Purdue University Indianapolis, 723 West Michigan Street, Indianapolis, IN, 46202, USA
[b] Indiana State Police Laboratory, 550 West 16th Street, Suite C, Indianapolis, IN, 46202, USA
[c] Forensic and Investigative Sciences, Indiana University Purdue University Indianapolis, 402 N. Blackford Street, Indianapolis, IN, 46202, USA

ARTICLE INFO

Keywords:
Forensic science
DNA
Mixtures
PowerPlex® fusion 6C
Allele counting

ABSTRACT

DNA mixtures are more frequently encountered in casework due to increased kit sensitivity, protocols with increased cycle number, and requests for low copy number DNA samples to be tested. Generally, the first step in mixture interpretation is determining the number of contributors, with the most common approach of maximum allele count. Although there are previous studies regarding the accuracy of this approach, none have evaluated the accuracy with the newly expanded U.S. core STR loci. In this work, 4,976,355 theoretical mixture combinations were generated with the PowerPlex® Fusion 6C system which includes 23 autosomal STR loci and three Y-STR loci. The number of contributors could be correctly assumed for 100% two-person and 99.99% three-person mixtures, whereas, four-, five-, and six-person mixtures were correctly assumed in 89.7%, 57.3%, and 7.8% of mixtures, respectively. Y-STR analysis showed the 3 Y-STR markers are only accurate for two-person male mixtures (96.7%). This work demonstrates that maximum allele count using the expanded U.S. core loci is not much improved from previous smaller panels, reiterating that this method is not as accurate beyond three contributors.

1. Background

Mixtures are a common challenge in DNA profile interpretation. DNA mixtures are more frequently encountered in forensic casework than in the earlier years of STR typing. This is mainly because of increased sensitivity of the commercially available genotyping kits and the opportunity to optimize PCR reactions for 'touch' or low copy number (LCN) DNA. One published study retroactively reviewed 1547 cases over four years (1997–2000). Of the 2424 samples from those cases, 163 (6.7%) showed a mixture profile, and only eight of the 163 (0.3%) samples were mixtures of more than two contributors [1]. A decade later, a survey study initiated by SWGDAM in 2008 collected case data from 14 laboratories on 4541 samples, where 45.2% showed a mixture profile, and 526 (11.6%) samples were mixtures of more than two contributors [2]. This survey was the basis of the 2010 SWGDAM DNA interpretation guidelines to focus on single source and two person mixture samples, although updated SWGDAM guidelines do include criteria for more than two contributors [3].

The first step in interpreting a DNA profile is identifying the presence of a mixture, or, a profile with more than one contributor. This is typically determined by analyzing the number of allelic peaks and peak height ratios, while considering stochastic effects, including stutter. According to SWGDAM guidelines, if one or more loci have 3 or more alleles present, excluding tri-allelic loci, then the sample is assumed to be a mixture [3]. The next logical step is determining the number of contributors in that mixture. This is a key step to the deconvolution of the mixture to assign genotypes to each individual present for providing statistical weight to the evidence. The most common approach for estimating the number of contributors is maximum allele count [4]. Maximum allele count is used to estimate the number of minimum contributors to the mixed sample by evaluating the locus that has the greatest number of allelic peaks [3], because a single individual should only have a maximum of two alleles at a locus. For example, if a locus has five allele peaks, there has to be a minimum of three contributors because for a two-person mixture, the expected maximum number of alleles is four.

There have been some previous studies to characterize the number of contributors according to maximum allele count. Paoletti et al. [5] generated conceptual three- and four- person mixtures from an FBI database which contained genotypes from the 13 core CODIS STR loci from 959 individuals. Based on maximum allele count, they found 3% of the 146,536,159 three-person mixtures could be mischaracterized as two-person mixtures, and that 76% of the 57,211,376 four-person

* Corresponding author at: Department of Biology, Indiana University-Purdue University Indianapolis, 723 West Michigan Street, Indianapolis, IN, 46202, USA.
E-mail addresses: gdembins@iupui.edu (G.M. Dembinski), CSobieralski@isp.in.gov (C. Sobieralski), cpicard@iupui.edu (C.J. Picard).

https://doi.org/10.1016/j.fsigen.2017.11.007
Received 23 August 2017; Received in revised form 6 November 2017; Accepted 12 November 2017
Available online 20 November 2017
1872-4973/

mixtures could be mischaracterized as two- or three-person mixtures [5]. Haned et al. [4] conducted simulations from published genotypes of individuals with 15 STR loci (including 13 CODIS loci) by generating 1000 mixtures comprised of between two to five contributors to compare maximum allele count with maximum likelihood, another method for determining number of contributors [4]. They concluded that mixtures of two or three contributors was greater than 90% for both methods, but with mixtures of 4 or 5 contributors, maximum likelihood yielded greater success rates. For example, correct mixtures for Caucasians with four and five contributors with maximum allele count was 34% and 2%, respectively, but with maximum likelihood, 77% and 64%, respectively [4].

There are three different mathematical models that can be used for mixture interpretation: binary, semi-continous, or continuous [6]. Binary statistical models (i.e., random match probability (RMP), likelihood ratio (LR), and combined probability of exclusion/inclusion (CPE/CPI)) are still very common in practice, however, they are limited to cases where at least one contributor can be deconvoluted from the mixture (RMP), or, the need for all alleles to be present for the loci to be used (CPI). They also require the analyst to assume the number of contributors in order to perform the statistic (with the exception of CPE/CPI) [3]. Furthermore, according to the recently released PCAST report, the CPE/CPI statistic was deemed inadequate and subjective [7]. Probabilistic genotyping methods and software (semi-continuous and continuous models) have been developed as an improved alternative to simpler binary practices for mixture interpretation, examples of programs include TrueAllele® [8], STRmix™[9], Lab Retriever [10], *forensim* R package [11], and NOCit, a program that estimates number of contributors as part of the PROVEDIt initiative [12]. However, the PCAST report points to the fact that although probabilistic genotyping methods are an improvement, further testing should be done to ensure the scientific validity on reliability and on the algorithms being implemented [7]. This need for further testing could be seen demonstrated in *New York v. Oral Hillary*, where two different probabilistic programs were used to evaluate the same mixture profile and two different conclusions resulted (TrueAllele® did not find a link to the defendant whereas STRMix™ could not rule out the defendant) [13]. Many laboratories are still implementing binary methods, including maximum allele count to determine number of contributors.

To date, there have not been any published studies to evaluate the maximum allele count method on the expanded U.S. core STR loci. The PowerPlex® Fusion 6C system (Promega Corp., Madison, WI) incorporates 27 loci which includes the expanded 20 U.S. CODIS core loci [14]. It was the objective of this work to evaluate how the maximum allele count method would determine number of contributors for theoretically generated combined two-, three-, four-, five-, and six-person mixtures (4,976,355 total mixture profiles) based on 236 unrelated genotypes using the PowerPlex® Fusion 6C kit (Promega Corp.).

2. Material and methods

Single source reference DNA profiles (N = 236) were amplified from non-related anonymous volunteers collected by the Indiana State Police Laboratory. Genotypes were generated using the PowerPlex® Fusion 6C System (Promega Corp.) using the BioMek NX^P and BioMek 3000 Automated Workstations under standard casework operating procedures of the Indiana State Police Laboratory [15]. The genotypes, each designated with a random number identifier, were entered into an electronic database using Microsoft Excel for theoretical mixture generation and analysis. There were 4,976,355 total mixture combinations generated.

A macro using Visual Basic in Microsoft Excel was used to generate all possible combinations of two- and three-person mixtures. The macro was also used to generate combinations for the four-, five-, and six-person mixtures, however, due to the large number of possible combinations and the limitation in number of rows possible in Excel (1,048,576 rows), only a random subset (generated using a random number function in Excel) of all possible combinations were analyzed (see Table 1). Two separate random sample sets were generated for the four-person combinations to ensure allele count distributions were representative of the whole set (Table 2). Statistics were performed in Microsoft Excel. Each set of mixtures were analyzed with the following defined parameters: (i) the minimum allele count is the count across all loci per profile that had the lowest number of alleles observed in at least one locus; (ii) the maximum allele count is the count across all loci per profile that had the highest number of alleles observed in at least one locus; and, (iii) the overall count is the frequency distribution of all allele counts across all loci of all possible *n* person profile combinations. As the kit also contains three Y-STR markers, a separate analysis of the Y-STRs was performed whereas the generated mixture combinations were filtered to analyze those between male mixtures only.

Table 1
Number of mixture combinations.

Number of Contributors	Number of mixtures generated (number of database samples used)
2 person	27,730 (236)
3 person	2,162,940 (236)
4 person	916,895 (70)[a]
5 person	962,598 (43)[a]
6 person	906,192 (32)[a]
TOTAL	4,976,355

[a] Subset of total possible combinations.

Table 2
Comparison of maximum allele count between two separately generated combinations of four-person mixtures (N = 916,895 each, $p > 0.5$).

4-person Mixture Combinations: Allelic Distribution		
Number of Alleles	Group 1	Group 2
1	5555	11204
2	533410	623641
3	3467181	3671196
4	6601424	6289826
5	5805377	5596867
6	3236480	3307998
7	1194533	1302849
8	244625	285004
TOTAL	21088585	21088585

3. Results and discussion

For all mixture combinations, the profiles were considered under ideal conditions (equal ratios, no stutter or artifacts, and all alleles were above the stochastic threshold). Therefore, allele count analysis was calculated based on the assumed presence of all possible allelic peaks from all individuals in the mixture. This does not reflect the possibilities of mixed ratios, stutter, or allele-dropout, which are not unexpected in casework mixtures [16,17].

3.1. Two-person mixtures

For the 27,730 two-person mixtures, the minimum allele count was two in 70% of mixtures, and one in the remaining 30% of profiles (Fig. 1a). Although it was possible to see loci with only one allele, which is more typical in single source profiles, a maximum allele count of four was 99.99% (Fig. 1b). There were four (0.01%) profiles that had a maximum of three allele in at least one locus. In either case, a minimum of two contributors would still be indicated (Fig. 1b). Based on maximum allele count, two-person mixtures could accurately be determined as having two contributors in all cases. SE33, D1S1656, and Penta E are among the loci that have the highest frequency of the