thermo scientific

Lead discovery

When you need to be at the forefront of discovery, the very highest data quality and comprehensive quantitative information are paramount. Advance your science and adapt to evolving research demands with the unprecedented performance and versatility of the Thermo Scientific™ Orbitrap Exploris™ GC 240 mass spectrometer. Get the right answer the first time and solve complex analytical challenges with easy access to the very highest-accuracy, information-rich GC-MS data, today and well into the future.

Delivering broader and deeper insights

Find out more at **thermofisher.com/OrbitrapExplorisGC240**

ThermoFisher SCIENTIFIC

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, March 2017, Vol. 62, No. 2
doi: 10.1111/1556-4029.13284
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

Lauren E. Alfonse,[1] *M.S.; Genesis Tejada*,[1] *M.S.; Harish Swaminathan*,[2] *Ph.D.; Desmond S. Lun*,[2,3] *Ph.D.; and Catherine M. Grgicak*,[1] *Ph.D.*

Inferring the Number of Contributors to Complex DNA Mixtures Using Three Methods: Exploring the Limits of Low-Template DNA Interpretation*

ABSTRACT: In forensic DNA casework, the interpretation of an evidentiary profile may be dependent upon the assumption on the number of individuals from whom the evidence arose. Three methods of inferring the number of contributors—NOCIt, maximum likelihood estimator, and maximum allele count, were evaluated using 100 test samples consisting of one to five contributors and 0.5–0.016 ng template DNA amplified with Identifiler® Plus and PowerPlex® 16 HS. Results indicate that NOCIt was the most accurate method of the three, requiring 0.07 ng template DNA from any one contributor to consistently estimate the true number of contributors. Additionally, NOCIt returned repeatable results for 91% of samples analyzed in quintuplicate, while 50 single-source standards proved sufficient to calibrate the software. The data indicate that computational methods that employ a quantitative, probabilistic approach provide improved accuracy and additional pertinent information such as the uncertainty associated with the inferred number of contributors.

KEYWORDS: forensic science, number of contributors, DNA mixture interpretation, maximum allele count, maximum likelihood estimator, NOCIt

Since variable number tandem repeats were first described by Gill et al. as hypervariable regions of DNA from which individualizing DNA "fingerprints" of forensic relevance may be derived, typing short tandem repeat (STR) markers has emerged as the most widely utilized strategy in human identification (1,2). Advances in laser and detection technologies and PCR chemistries have made possible the detection of small quantities of DNA. Commonly referred to as low-template DNA (LTDNA), samples that exhibit low RFU peaks are now frequently encountered and interpreted in forensic laboratories (3–5). The ability to successfully acquire electropherogram signal from minimal quantities of DNA has led to an increase in the demand to process multisource mixture samples (6).

The presence of artifacts and stochastic phenomena in an electropherogram can complicate the discernment of contributor genotypes in a mixture. For example, the detection of PCR products one repeat shorter (reverse stutter) or one repeat longer (forward stutter) than an allele, the detection of extraneous peaks (drop-in), and the failure to detect an allele (drop-out) can all complicate DNA interpretation (5). Additionally, because most commercially available STR amplification kits are optimized for use with 0.5–1.5 ng of DNA, amplification of <0.25 ng leads to increased signal variation, such as peak height imbalance, which can further confound mixture interpretation (6,7).

In response to the increase in complex LTDNA mixtures in forensic casework, several software packages have been developed to supplement or replace manual interpretation (8–15); however, many currently available DNA interpretation software require the user to specify a value for the number of contributors to the sample prior to analysis (8,10,12–15). Despite these advances, the interpretation of complex, LTDNA mixtures continue to pose a challenge, as the capacity to correctly deconvolve and statistically assess such mixtures is dependent upon the accuracy of the assumption on the number of contributors. If this assumption is incorrect, downstream interpretation, which includes genotype deconvolution and statistical interpretation, may be affected (16).

Threshold-based methods of determining the number of contributors include the maximum allele count (MAC) and the maximum likelihood estimator (MLE). Using MAC, the minimum number of contributors to a sample is determined by evaluating which locus in a profile exhibits the greatest number of obligate alleles (after removal of artifacts), dividing the total number of alleles at that locus by two, and rounding up to the nearest

[1]Biomedical Forensic Sciences, Boston University School of Medicine, Boston, MA.
[2]Center for Computational and Integrative Biology, Rutgers University, Camden, NJ.
[3]School of Information Technology and Mathematical Sciences, University of South Australia, Adelaide SA, Australia.

*Partially supported by 2011-DN-BX-K558 and 2012-DN-BX-K050 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice.
The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect those of the Department of Justice.

Received 24 Sept. 2015; and in revised form 11 June 2016; accepted 25 June 2016.

whole number. Although the analyst may consider peak heights in evaluating the number of contributors determined by MAC, this approach is problematic when used with complex, higher-order mixtures which commonly exhibit allele sharing; when different contributors present one or more of the same alleles at a locus, the number of observed alleles may be an underrepresentation of the true number of contributors to the sample (17–19). Improved accuracies for higher-order mixtures were reported using the MLE function in *forensim* —an R package designed for the statistical interpretation of forensic DNA mixtures (20). The likelihood estimator functions calculate the likelihood of observing a set of alleles in an evidentiary profile, contingent on the number of contributors to the sample, based on the frequency of the observed alleles in a given population. Utilizing simulated mixtures, Haned et al. (20) determined MLE to yield accuracy rates 2- to 15-fold higher for four- and five-contributor mixtures compared to MAC. Other approaches, all of which omit use of quantitative data present in an electropherogram, have also been proposed (21–23).

This work seeks to evaluate NOCIt—a continuous computational tool which calculates the *a posteriori* probability (APP) on the number of contributors to a DNA sample by examining all signal, including baseline noise (24). Because NOCIt utilizes Monte Carlo sampling, phase I of this study explores the precision and repeatability of NOCIt analysis. In phase II, the minimum number of standards needed to calibrate the software is determined. Finally, in phase III, three methods of inferring the number of contributors are evaluated by examining 20 single-source and 80 multicontributor test samples prepared using Identifiler® Plus and PowerPlex® 16 HS protocols. The three methods explored include MAC, MLE, and NOCIt, chosen because they range in computational sophistication from a counting method (MAC) to a continuous system which does not utilize any thresholds (NOCIt). The effect of sample complexity on the ability to correctly determine the number of contributors using MAC, MLE, and NOCIt is then considered using real samples consisting of one to five contributors and 0.5–0.016 ng template DNA.

Methods

Description of NOCIt

NOCIt is an interpretation tool designed to evaluate the number of contributors that gave rise to an unknown or evidentiary DNA sample by interpreting all data generated by autosomal STR signal, including baseline noise. Described extensively in (24), NOCIt, in short, provides a probability distribution on the number of contributors. Assuming a uniform *a priori* distribution between zero and five contributors, NOCIt computes the likelihood that n unrelated individuals contributed to a sample via a Monte Carlo sampling process during which genotypes for the n contributors are chosen based on the allele frequencies provided, and the ratio of the n contributors to the mixture is selected from a uniform distribution. Modeling of baseline noise, reverse and forward stutter proportions, stutter and allele drop-out rates, and allele heights is also performed. The likelihood of observing the peak heights at each locus given the genotypes of the contributors, the mixture ratio, and the template DNA mass is computed using peak height information generated by examining single-source calibration standards of known genotype. The output is presented as a probability distribution over zero to five contributors, providing the *a posteriori* probability that n contributors gave rise to the evidence sample. Therefore, unlike counting methods, NOCIt provides information not only on the most probable value (i.e., the number of contributors corresponding to the maximum *a posteriori* probability, MAP), but also the probabilities associated with other values (24).

Data Generation

Single-source data were obtained by extracting DNA from whole blood or proficiency test samples purchased from various manufacturers using phenol/chloroform purification and alcohol precipitation. The extracts were quantified using Quantifiler® Duo (Life Technologies, Carlsbad, CA) on the Applied Biosystems® 7500 (Applied Biosystems, Foster City, CA) using the manufacturer's recommended thermal cycling protocol and a validated, universal calibration curve (25,26). The extracts were amplified on the GeneAmp® PCR Amplification System 9700 with a gold sample block (Applied Biosystems) using AmpF*l*str® Identifiler® Plus (Life Technologies) following the manufacturer's recommended protocol (29 cycles) at the following target masses: 0.5, 0.25, 0.125, 0.063, 0.047, 0.031, 0.016, 0.008 ng (27). Fifty extracts were re-quantified and amplified on the GeneAmp® PCR Amplification System 9700 with a gold sample block using the PowerPlex® 16 HS System (32 cycles), following the manufacturer's recommended protocol (25,28).

A portion of the single-source extracts and corresponding quantitation values obtained above were used to prepare mixture samples, which ranged from two to five contributors and 0.5–0.008 ng total template DNA. The quantitation values were used to calculate the appropriate volume of each single-source extract to combine based on the desired number of contributors to the mixture and the contributor ratio. Where necessary, dilutions were prepared in TE buffer. After the single-source extracts were combined, the mixtures were re-quantified using the aforementioned procedure and amplified using the same protocol described for single-source samples.

All amplicons were injected for 10 sec at 3 kV on the Applied Biosystems® 3130 Genetic Analyzer (Applied Biosystems), and resultant electropherograms were analyzed with GeneMapper® ID-X v1.1.1 (Applied Biosystems) at an analytical threshold of 1 RFU. Artifacts such as pull-up, complex pull-up, and minus A were manually removed. Pull-up was defined as a peak which appears in the same position (±0.3 bases) as an allelic peak in another color channel and has a peak height of 5% or less of the allelic peak. Complex pull-up was defined as a peak with a plateau-like shape located between two adjacent allelic peaks in a different color channel. Minus A was defined as a peak one base shorter in size (± 0.3 bases) than an allelic peak. There were no height restrictions for the complex pull-up and minus A artifacts. Additionally, instances of dissociated dye, defined as a blob-shaped peak appearing in the same location and color channel across multiple samples, were manually removed from mixture samples. The filtered genotype table for each sample was exported from GeneMapper® ID-X as a CSV file containing the allele, size, and height for all peaks.

I. NOCIt Precision and Repeatability

Two single-source and 126 multicontributor samples amplified with Identifiler® Plus were run in quintuplicate on NOCIt, resulting in a total of 640 NOCIt runs (Table 1; Table S1). All runs utilized an allele frequency file created from the Caucasian allele

TABLE 1—*The samples tested in each of the three phases of this study.*

	NOC	1	2	3	4	5	Total
I. NOCIt precision and repeatability	Number of test samples	2	4	36	28	58	128
	Total template mass (ng)	0.25	0.5–0.016	0.5–0.008	0.25–0.008	0.25–0.008	
	Contributor ratios	N/A	1:2	1:1:1	1:1:1:1	1:1:1:1:1	
			1:4	1:2:1	1:1:2:1	1:1:2:1:1	
			1:9	1:4:1	1:1:4:1	1:1:4:1:1	
			1:19	1:9:1	1:1:9:1	1:1:9:1:1	
				1:2:2	1:2:2:1	1:1:2:2:1	
				1:4:4	1:4:4:1	1:1:4:4:1	
				1:9:9	1:9:9:1	1:1:9:9:1	
						1:2:2:2:1	
						1:4:4:4:1	
						1:9:9:9:1	
II. NOCIt calibration	Number of test samples	7	8	7	3	5	30
	Total template mass (ng)	0.25–0.008	0.063–0.008	0.063	0.25	0.25	
	Contributor ratios	N/A	1:1	1:1:1	1:1:1:1	1:1:1:1:1	
			1:2	1:2:1	1:1:2:1	1:1:2:1:1	
			1:4	1:4:1	1:4:4:1	1:1:2:2:1	
			1:9	1:9:1		1:2:2:2:1	
			1:19	1:2:2		1:1:4:4:1	
				1:4:4			
				1:9:9			
III. Performance of NOCIt, MAC, and MLE	Number of test samples	20	20	20	20	20	100
	Total template mass (ng)	0.25–0.016	0.25–0.016	0.5–0.05	0.25–0.063	0.25–0.063	
	Contributor ratios	N/A	1:1	1:1:1	1:1:1:1	1:1:1:1:1	
			1:2	1:2:1	1:1:2:1	1:1:2:1:1	
			1:4	1:4:1	1:1:4:1	1:1:4:1:1	
			1:9	1:9:1	1:1:9:1	1:1:9:1:1	
			1:19	1:2:2	1:2:2:1	1:1:2:2:1	
				1:4:4	1:4:4:1	1:1:4:4:1	
				1:9:9	1:9:9:1	1:1:9:9:1	
						1:2:2:2:1	
						1:4:4:4:1	
						1:9:9:9:1	

frequencies provided in the Identifiler® Plus User's Guide and a calibration file containing 95 single-source calibration standards each amplified with Identifiler® Plus at the following eight target masses: 0.5, 0.25, 0.125, 0.063, 0.047, 0.031, 0.016, and 0.008 ng. In the current and proceeding phases of this study, all NOCIt files were formatted in accordance with the guidelines set forth in (29). Additionally, when single-source test samples were evaluated on NOCIt, they were excluded from the calibration data.

II. NOCIt Calibration

To determine the minimum number of standards needed to calibrate NOCIt, seven calibration files, containing 95, 80, 65, 50, 35, 20, and 15 standards, were created. The calibration files were created such that each subsequent file contained fewer standards than the previous. For example, the first calibration file contained 95 standards; the second calibration file contained 80 of the 95 standards; and the third contained 65 of the 80 standards. The calibration standards were each amplified with Identifiler® Plus at the following seven target masses: 0.25, 0.125, 0.063, 0.047, 0.031, 0.016, and 0.008 ng. The full mass range was represented for every standard in the calibration files, because the calibration data must capture different levels of drop-in, drop-out, and other stochastic phenomena to enable the various NOCIt models to function optimally. Thirty test samples, ranging from one to five contributors, were then analyzed once each on NOCIt with the set of seven calibration files, utilizing the Identifiler® Plus Caucasian allele frequency file (Table 1; Table S1).

III. Performance of NOCIt, MAC, and MLE

A total of 100 test samples ranging from one to five contributors amplified with both Identifiler® Plus and PowerPlex® 16 HS were analyzed once each using NOCIt, MAC, and MLE (Table 1; Table S1). For NOCIt analysis, calibration files prepared for use with each kit contained 50 single-source standards each amplified at the following six target masses: 0.25, 0.125, 0.063, 0.031, 0.016, and 0.008 ng. Caucasian allele frequencies provided by each manufacturer were utilized (27,28).

During analysis with MAC and MLE, locus-specific reverse stutter filters were applied to all samples in accordance with values provided by the manufacturers (27,28). Additionally, all samples were analyzed at 50 RFU—a commonly applied analytical threshold (30). This analytical threshold has also been shown to provide a reasonable level of protection against the false detection of noise for both amplification kits (31).

Using MAC, the minimum number of contributors to each sample was determined by evaluating which locus in the profile exhibited the greatest number of alleles, dividing the total number of alleles at that locus by two, and rounding up to the nearest whole number. Analysis with MLE was achieved by installing and opening the FORENSIM package in R v3.0.3 (The R Foundation for Statistical Computing, Vienna, Austria). Caucasian allele frequency datasets created from values provided by the manufacturers were converted to tabfreq objects using the tabfreq constructor, and each test sample, previously saved as a CSV, was converted to a simumix object using the simumix constructor. A probability distribution on the number of contributors for each sample was obtained by executing the *lik* function

which determines the likelihood of observing the sample given *n* contributors, where the value of *n* was varied from one to six (32).

Results and Discussion

A list of abbreviations is provided in (Table 2).

A Representative NOCIt Result

Figure 1 depicts a representative NOCIt result obtained from analysis of two test samples: one simple and one complex mixture. The former consists of 0.25 ng template DNA from one contributor. As shown in Fig. 1*A*, the MAP (0.99) indicates that there is a high probability that the sample arose from one contributor. This corresponds to the known number of contributors, indicating that NOCIt correctly evaluated the sample and determined that it is highly probable that the sample was generated from the amplification of one individual's DNA. For reference, a portion of the electropherogram for this sample is shown in Fig. 1*B*.

In contrast, analysis of complex mixtures often suggests that there is some uncertainty in determining the number of contributors. For example, Fig. 1*C* shows the resultant probability distribution obtained from analysis of a five-contributor sample consisting of 0.25 ng total template DNA at a 1:1:9:1:1 contributor ratio; a portion of the electropherogram for this sample is shown in Fig. 1*D*. Although the MAP (0.68) corresponds to four contributors, the distribution indicates that a probability of 0.32 was assigned to the correct, known number of contributors (i.e., five). The interpretation is such that there is a probability of 0.68 that this evidence would be observed if it originated from four contributors, as well as a 0.32 probability of observing the evidence if it arose from five contributors. Some uncertainty in the probability distribution is unsurprising given that the mass of each minor contributor is approximately 0.02 ng, and drop-out and allele sharing are likely to be observed. In short, the NOCIt output indicates that it is probable that the sample arose from four contributors; however, due to the complexity of the sample, other values may explain the evidence and should be considered.

I. NOCIt Precision and Repeatability

The precision of NOCIt was examined by recording the number of times the NOCIt MAP estimate corresponded to the same number of contributors, *n*, or a different *n* between the quintuplicate runs of each sample. Ideally, a given sample, when tested multiple times on NOCIt, would be estimated as the same number of contributors each time, with similar probabilities for each *n*. Figure 2 shows the results for all 128 samples tested, where the dark gray bars indicate the number of samples for which the MAP estimate corresponded to the same *n* for all five runs (a precise analysis), and the light gray bars indicate the number of samples for which the MAP estimate corresponded to varying values of *n* between the quintuplicate runs (an imprecise analysis). For 91% of the samples tested on NOCIt, the MAP estimate

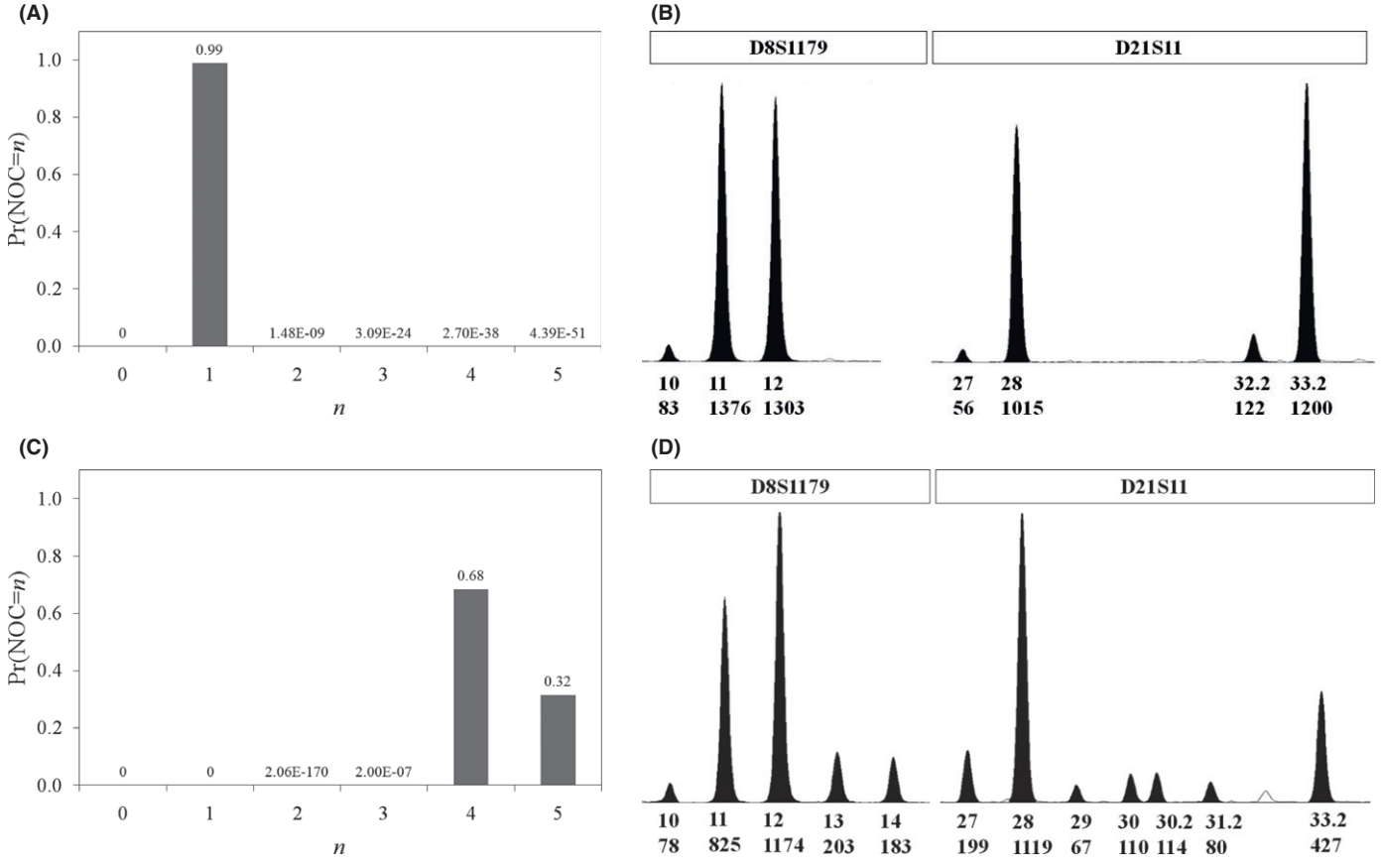

FIG. 1—*A graphical view of the NOCIt output, which presents the probability of observing a sample given n contributors, for (A) a one-contributor sample containing 0.25 ng template DNA and (C) a five-contributor sample containing 0.25 ng total template DNA at a 1:1:9:1:1 contributor ratio. Representative electropherograms, showing loci D8S1179 and D21S11, for (B) the one-contributor sample and (D) the five-contributor sample, indicate repeat numbers and peak heights for all allele and reverse stutter peaks called above 50 RFU.*

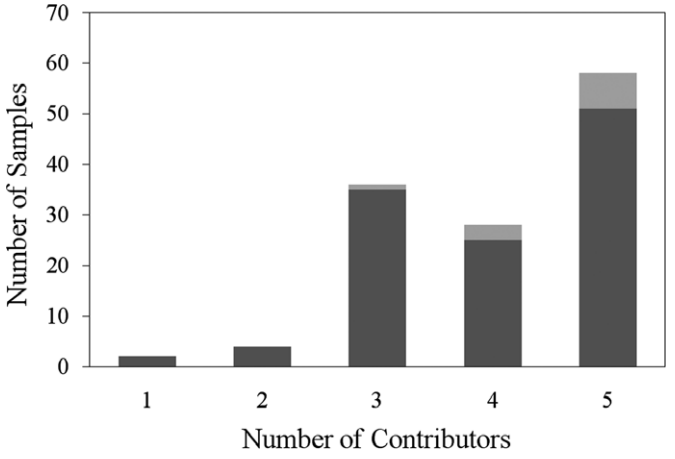

FIG. 2—*The number of samples for which the $NOCIt_{MAP}$ estimate corresponded to the same value for the number of contributors (■) or different values for the number of contributors (■) for quintuplicate runs of 128 samples.*

corresponded to the same value of *n* for all five replicates; however, Fig. 2 shows that imprecise estimates of the number of contributors occurred more frequently with five-contributor mixtures compared to lower-order samples.

The repeatability results are explored in Fig. 3, where the median MAP is plotted against the MAP range for the quintuplicate runs of each sample. For samples in which the median MAP consistently corresponded to the same value of *n* for all five replicates, the median MAP was obtained by calculating the median of all five MAP values obtained across the quintuplicate runs, while the MAP range was determined by subtracting the smallest MAP value from the largest MAP value obtained across the five runs. However, for 9% of samples tested (i.e., 11 of 128), quintuplicate NOCIt runs suggested that the data may be explained by two probable values for *n* (Table 3). As shown in Table 3, when the NOCIt MAP estimate indicated an incorrect value for *n*, this value was most often an underestimate of the known number of contributors, likely due to the presence of one or more extreme minor contributors. For example, during NOCIt analysis of a four-contributor sample (0.25 ng total template mass), for four of the five replicate runs, the MAP estimate corresponded to four contributors, and for one of the replicates, the MAP estimate corresponded to three contributors. For this sample and in similar instances, the median MAP and MAP range were determined by considering the value of *n* to which the majority of MAP values obtained corresponded. Given the results of the four-contributor analysis outlined above, the median of the probability estimate and the range of the probabilities obtained for $n = 4$ across the five runs were computed.

As shown in Fig. 3, the highest MAP NOCIt can assign to an *n* is 1, and the data indicate that the majority (i.e., 101 of 128) of the median MAP values occurred between 0.9 and 0.99 with a MAP range of <0.2. Thirteen samples yielded a MAP range >0.2; in each of these samples, the mass of the minor contributor(s) is <0.05 ng, indicating that the presence of extreme minor contributors may influence the amount of variation obtained across replicate runs. Further, it was observed that when the median MAP is less than approximately 0.85, the MAP range tends to be more variable, suggesting that the MAP output obtained for a sample may be used to indicate whether the sample will consistently produce the same number of contributors between replicate runs. That is, when inferring the number of

TABLE 2—*Abbreviations.*

Abbreviation	Definition
APP	*A posteriori* probability
MAP	Maximum *a posteriori* probability
$MLE_{1\%}$	≥1% probability generated by MLE
MLE_{MAP}	Maximum *a posteriori* probability generated by MLE
NOC	Number of contributors
$NOCIt_{1\%}$	≥1% probability generated by NOCIt
$NOCIt_{MAP}$	Maximum *a posteriori* probability generated by NOCIt
NOC_{known}	Known number of contributors

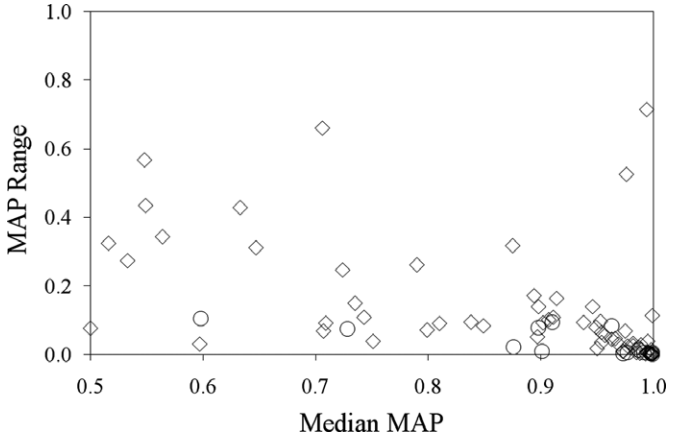

FIG. 3—*The MAP range versus the median MAP for quintuplicate NOCIt runs of 128 samples with minor contributor mass <0.05 ng (◇) and greater than or equal to 0.05 ng (○).*

TABLE 3—*Information regarding the 11 samples for which the $NOCIt_{MAP}$ estimate correlated to two different values of n during quintuplicate NOCIt runs in Phase I. The known number of contributors, contributor ratio, and total template mass is specified for each sample. The number of runs (out of five) in which $NOCIt_{MAP}$ corresponded to a given n is shown for each sample, where one gray box represents one run.*

Known NOC	Contributor Ratio	Total Template Mass (ng)	$NOCIt_{MAP}$	Number of Runs 1 2 3 4 5
3	1:4:1	0.25	$n = 3$ $n = 4$	■ ■■■■
4	1:4:4:1	0.063	$n = 2$ $n = 3$	■ ■■■■
4	1:2:2:1	0.125	$n = 3$ $n = 4$	■■■■ ■
4	1:9:9:1	0.25	$n = 3$ $n = 4$	■ ■■■■
5	1:1:2:1:1	0.063	$n = 3$ $n = 4$	■■ ■■■
5	1:1:4:1:1	0.063	$n = 3$ $n = 4$	■■■ ■■
5	1:9:9:9:1	0.063	$n = 3$ $n = 4$	■■■■ ■
5	1:1:2:2:1	0.125	$n = 4$ $n = 5$	■■■ ■■
5	1:1:2:2:1	0.25	$n = 4$ $n = 5$	■ ■■■■
5	1:1:4:4:1	0.25	$n = 4$ $n = 5$	■■■ ■■
5	1:4:4:4:1	0.25	$n = 4$ $n = 5$	■■■ ■■

contributors to a sample, in instances in which the MAP is <0.9, it may be of value to run the sample multiple times.

II. NOCIt Calibration

It was empirically determined that 50 single-source standards are sufficient to calibrate NOCIt. Figure 4 depicts the proportion of correct NOCIt estimates versus the proportion of incorrect estimates for each calibration file, where the number of contributors corresponding to the maximum APP estimated by NOCIt ($NOCIt_{MAP}$) was compared to the known number of contributors (NOC_{known}). A correct estimate was defined as $NOCIt_{MAP} = NOC_{known}$, while an incorrect estimate was defined as $NOCIt_{MAP} \neq NOC_{known}$. The correct estimate proportion was calculated by dividing the number of correct estimates by the total number of samples tested; the incorrect estimate proportion was computed similarly. As shown, the calibration files containing 95–50 standards amplified at seven target masses per standard all resulted in correct and incorrect estimate proportions of 100% and 0%, respectively. When the calibration file contained fewer than 50 standards, the correct estimate proportion decreased and the incorrect estimate proportion increased, resulting in reduced accuracy values. It should be noted that the calibration file containing 35 standards resulted in one error: Analysis of a one-contributor sample (0.25 ng template mass) resulted in a $NOCIt_{MAP}$ estimate of two contributors, possibly due to an incidence of forward stutter in excess of 2% (RFU N+1 peak/RFU N peak). However, the calibration files containing 20 and 15 standards resulted in more errors, particularly when analyzed with higher-order mixtures.

Additionally, it was observed that certain functionalities of NOCIt, such as the ability to model forward stutter at some loci, became compromised when 20 or fewer calibration standards were used. When 15 or fewer calibration standards were used to train NOCIt, some loci were excluded from probability calculations. Because the models for stutter, peak height, and drop-out utilize information from only heterozygous loci in the calibration data, the empirical determination of the minimum number of standards needed to calibrate NOCIt is inherently dependent upon the constituent genotypes of the calibration standards (29). The extent of information afforded to the models is further limited by the allelic patterns present in the calibration standards.

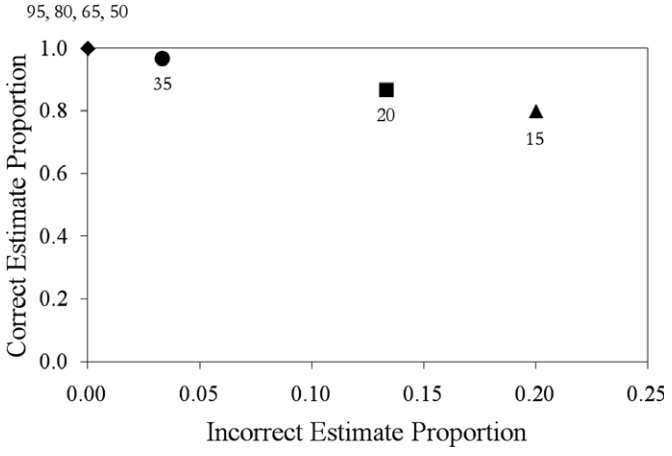

FIG. 4—*The proportion of correct estimates versus the proportion of incorrect NOCIt estimates for each calibration file. The number of standards in the calibration file is specified beside each point.*

For example, a standard with a known genotype of 11, 12 at a given locus, although heterozygous, would be excluded from consideration by the stutter models, as the RFU of the 11 peak is a combination of true allele and reverse stutter from the 12 allele, while the RFU of the 12 peak is a combination of true allele and forward stutter from the 11 allele. This phenomenon may be exacerbated if the calibration data contain a limited number of standards, some of which are low template and exhibit drop-out, leading to increased loss of meaningful information for the models.

Although 50 single-source standards were found to be sufficient to calibrate NOCIt in this set of experiments, if a different set of 50 standards with increased levels of homozygosity amplified at a reduced range of template masses is utilized, this dataset may provide insufficient information for the models and may thus yield inaccurate results. Therefore, the 50 calibration standard minimum is not intended as a blanket threshold, and it is advisable to repeat an empirical determination if calibration standards of new and/or different genotypes are utilized. Alternatively, 50 single-source standards with high numbers of heterozygous loci may be used.

III. Performance of NOCIt, MAC, and MLE

The accuracies of all three analysis methods for both kits are depicted in Fig. 5. Two methods of assessing accuracy were computed: the first—the accuracy of the $NOCIt_{MAP}$ estimate was computed for a given n by adding the number of times $NOCIt_{MAP} = NOC_{known}$ and dividing by the total number of samples tested. However, because NOCIt provides a probability distribution, in the analysis of highly complex mixtures, NOCIt results may indicate that more than one value for the number of contributors is possible. For example, in instances in which $NOCIt_{MAP} \neq NOC_{known}$, the NOCIt probability estimate for NOC_{known} is often significant and >1%; if a given n is assigned a probability of at least 1%, that n may represent the true value for the number of contributors and, as such, should be considered one of potentially multiple viable values for the number of contributors during downstream interpretation (24). Thus, the second form of accuracy—that of the 1% probability estimate ($NOCIt_{1\%}$), was computed for a given n by adding the number of times NOCIt estimated at least a 1% probability for NOC_{known} and dividing by the total number of samples tested. The number of contributors was also determined using the MAC and MLE methods and compared to NOC_{known}, and the accuracy of those results is shown in Fig. 5. The method for calculating the accuracy of the MLE maximum APP (MLE_{MAP}) and MLE 1% probability ($MLE_{1\%}$) estimates was the same as described above for the accuracy of $NOCIt_{MAP}$ and $NOCIt_{1\%}$, respectively. Additionally, because the likelihood estimator functions estimate an infinite number of contributors while the NOCIt probability distribution output is limited to the discrete interval 0–5, any five-contributor samples that were over-estimated as $n = 6$ with MLE were considered "correct" outputs (i.e., "accurate" results).

As shown in Fig. 5, NOCIt resulted in the highest accuracies for all one-, two-, three-, and four-contributor samples tested. MAC and MLE resulted in similar accuracies for one-, two-, and three-contributor samples, but for four- and five-contributor samples, MLE outperformed MAC. The accuracies obtained for five-contributor samples analyzed with NOCIt and MLE were similar, while all five-contributor samples analyzed with MAC were incorrectly estimated as having originated from four or fewer contributors. Specifically, the data obtained using the

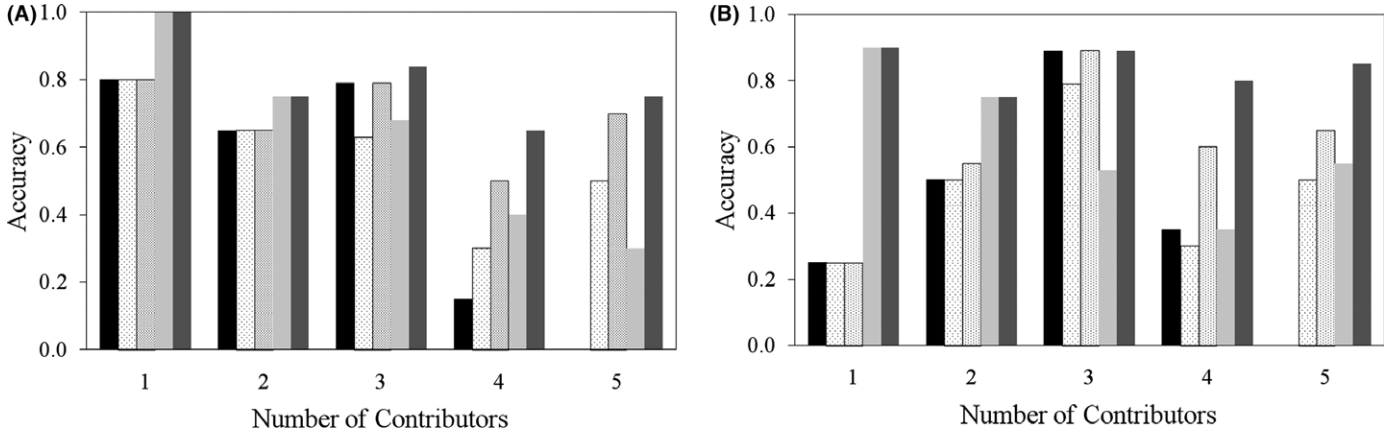

FIG. 5—*The accuracies of MAC (■), MLE_{MAP} (⬚), $MLE_{1\%}$ (▦), $NOCIt_{MAP}$ (▩), and $NOCIt_{1\%}$ (▩) versus the known number of contributors for samples amplified with (A) Identifiler® Plus and (B) PowerPlex® 16 HS.*

Identifiler® Plus kit, shown in Fig. 5*A*, indicate that the accuracy of $NOCIt_{MAP}$, MLE_{MAP}, and MAC generally decreased as the number of contributors increased; however, the accuracy of $NOCIt_{1\%}$ remained in excess of 65% for all samples tested, regardless of the number of contributors.

Comparing Fig. 5*A,B*, it is evident that NOCIt performed similarly with both amplification kits tested; however, the accuracies of MAC and MLE were lower with one- and two-contributor samples amplified with PowerPlex® 16 HS compared to Identifiler® Plus. The decreased accuracy observed with MLE and MAC is thought to be the result of above-threshold *N*-2 (i.e., double-back) stutter peaks observed in PowerPlex® 16 HS profiles which caused overestimates of the number of contributors to lower-order (i.e., one- and two-contributor) samples. The impact of *N*-2 stutter has recently been studied, and the findings of Mönich et al. (31) suggest that *N*-2 and *N*+2 stutter impact PowerPlex® 16 HS noise data in a significant way, particularly at larger concentrations of DNA. Thus, the misclassification of instances of *N*-2 stutter as allelic peaks tends to result in an overestimate of the number of contributors. This highlights the need to incorporate computational systems that account for allele drop-in or *N*-2 stutter, as binary methods seem incapable of properly evaluating all nuances associated with low-template DNA interpretation.

Generally, the accuracy results obtained are concordant with those presented in the literature. Utilizing simulated mixtures, Haned et al. (20) determined MLE to yield accuracy rates 2- to 15-fold higher for four- and five-contributor mixtures compared to MAC. Further, for a four-contributor mixture amplified and analyzed in quadruplicate, Benschop et al. (33) found that MLE consistently yielded underestimates of the number of contributors unless a consensus profiling approach was utilized, where alleles present in at least two of the four replicates were assigned to the consensus profile. Finally, Swaminathan et al. (24) found that the accuracy of MLE exceeded that of MAC for four- and five-contributor samples, while NOCIt resulted in the highest accuracies across all samples tested.

For all methods investigated, it was observed that the accuracy of the number of contributors estimate decreases as the total template mass decreases and as the ratio between any two contributors becomes more discrepant. Because the actual template mass of each contributor may play a substantive role in determining the number of contributors to complex mixtures, the accuracies of all methods were further explored in terms of the mass of the minor contributor for all samples tested (Fig. 6). The accuracies were calculated in the same manner stated previously, and the mass of the minor contributor was calculated for each sample by dividing the total template mass by the sum of the constituent parts of the known contributor ratio and rounding to the nearest two hundredth.

The results in Fig. 6 display accuracy versus mass of the minor contributor. The accuracies of $NOCIt_{MAP}$, MLE_{MAP}, and MAC are shown in (A) and (C) for samples amplified with Identifiler® Plus and PowerPlex® 16 HS, respectively, while the accuracies of $NOCIt_{1\%}$ and $MLE_{1\%}$ for samples amplified with Identifiler® Plus and PowerPlex® 16 HS are plotted in (B) and (D), respectively. Overall, the results indicate that $NOCIt_{1\%}$ resulted in the highest accuracy for the number of contributor estimations of all methods, across all minor contributor template masses tested. $NOCIt_{1\%}$ generally outperformed $MLE_{1\%}$, while MLE_{MAP} and MAC performed similarly. The accuracy of the $NOCIt_{1\%}$ estimate always equaled or exceeded that of $NOCIt_{MAP}$; the same trend was observed with $MLE_{1\%}$ and MLE_{MAP}. It is important to note that using the 1% probability threshold for both NOCIt and MLE does not yield one correct value for the number of contributors, but rather a range of values in which the true number of contributors may lie. Thus, utilization of the 1% probability accuracies of NOCIt and MLE suggests that consideration of more than one probable value for the number of contributors during downstream interpretation may be warranted.

Figure 6 shows that the accuracy of $NOCIt_{MAP}$ reaches 100% when the mass of the minor contributor is 0.07 ng, across all samples tested, irrespective of amplification kit. These results suggest that to be consistently detected by NOCIt, all of the contributors tested were required to have the equivalent of approximately 10 cells' worth of DNA present during amplification with either kit. At the same minor contributor mass, the accuracy of $MLE_{1\%}$ and MAC for samples amplified with Identifiler® Plus was 91% and 82%, respectively. Similarly, the accuracy of $MLE_{1\%}$ and MAC was 64% and 55%, respectively, for samples amplified with PowerPlex® 16 HS.

Cumulatively, the results in Fig. 6 indicate that, depending on analysis method, two phenomena may occur: (i) As the mass of the minor contributor decreases below 0.05 ng, the accuracy of the number of contributors estimate decreases due to drop-out which results in a tendency to underestimate, and (ii) as the mass of the minor contributor increases above 0.09 ng, accuracy similarly decreases due to a tendency to overestimate the number of contributors. The latter occurs because the total template mass of the mixture becomes large when the mass of each contributor

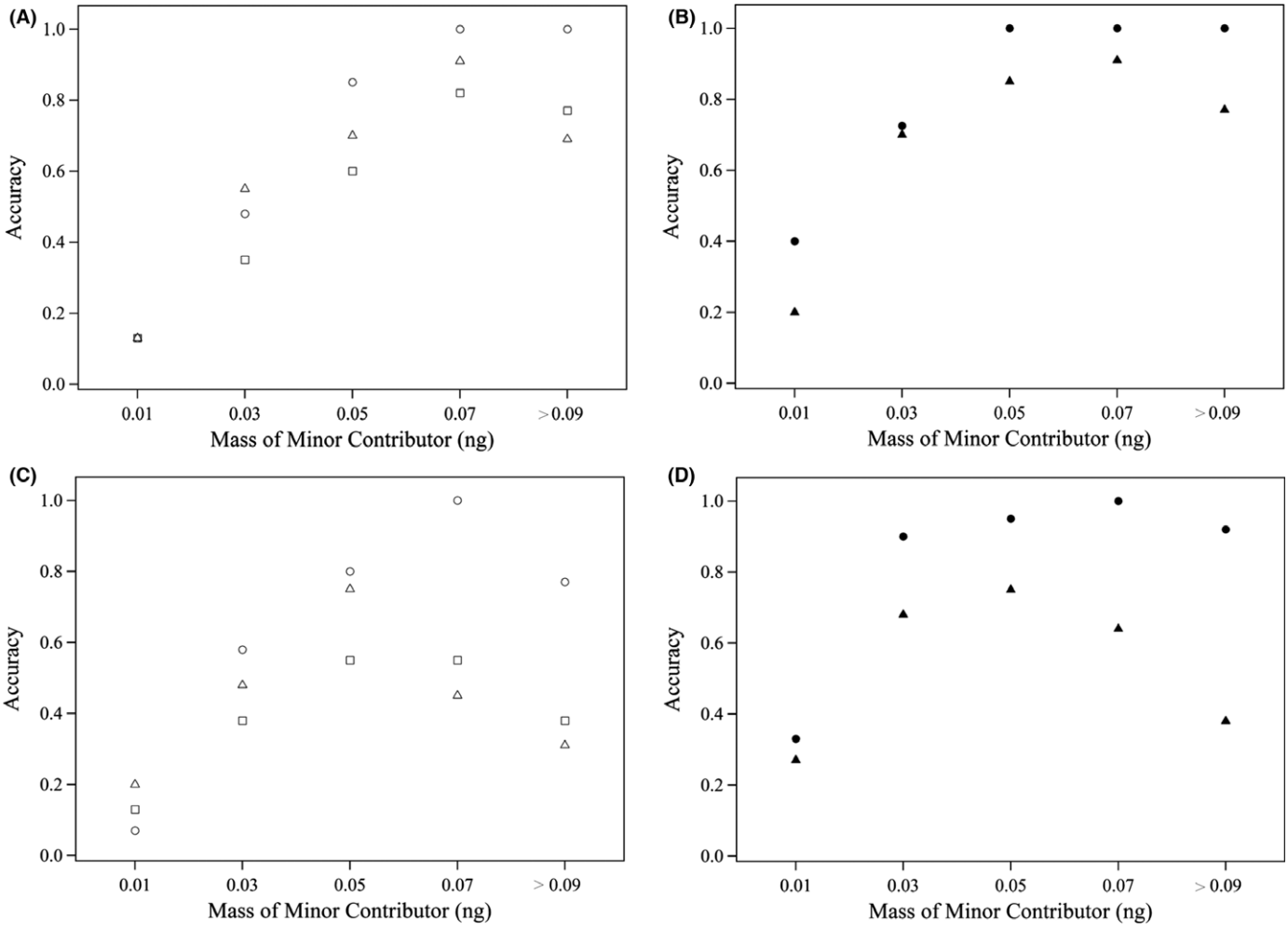

FIG. 6—*The accuracies of MAC (□), MLE_{MAP} (Δ), and $NOCIt_{MAP}$ (○) versus the mass of the minor contributor for samples amplified with (A) Identifiler® Plus and (C) PowerPlex® 16 HS, and the accuracies of $MLE_{1\%}$ (▲) and $NOCIt_{1\%}$ (●) versus the mass of the minor contributor for samples amplified with (B) Identifiler® Plus and (D) PowerPlex® 16 HS. Data displayed depict the composite accuracy values for all one-, two-, three-, four-, and five-contributor samples tested.*

is at least 0.09 ng, resulting in high instances of forward stutter and drop-in. Although the accuracies of all three methods decreased when the mass of the minor contributor was <0.05 ng, NOCIt was the only method to exhibit a reasonable tolerance to higher masses. Unlike MLE and MAC, NOCIt did not show a significant decrease in accuracy when the mass of the minor contributor exceeded 0.09 ng (corresponding to a total template mass of 0.125–0.5 ng). Finally, it is noted that when the mass of the minor contributor is 0.01 ng—the equivalent of approximately two cells' worth of DNA, the accuracy of the $NOCIt_{1\%}$ estimate for samples amplified with Identifiler® Plus and PowerPlex® 16 HS is 40% and 33%, respectively, indicating that the ability to detect and discern the presence of a contributor is a critical, underlying factor affecting the accuracy of DNA interpretation. Cumulatively, these results suggest that analysis of signal from minor components may be enhanced through the implementation of a DNA processing strategy which utilizes systems, such as NOCIt, that are unencumbered by thresholds.

Conclusion

The results obtained herein support the utility of a probabilistic approach to inferring the number of contributors to complex, multicontributor DNA samples. Phases I and II of this study indicated that NOCIt returned repeatable results for 91% of samples analyzed in quintuplicate, while 50 single-source standards, each amplified at seven target masses, proved to be sufficient to calibrate the software. Additionally, the results obtained in phase III found that the ability to correctly estimate the number of contributors to forensic DNA mixtures is affected by analysis method, where NOCIt was found to outperform the two existing methods investigated. The effect of analysis method was further explored using two different amplification kits; although NOCIt proved to be a robust system and performed similarly with both kits, a decrease in the accuracy of MAC and MLE was observed with one- and two-contributor samples amplified with PowerPlex® 16 HS compared to Identifiler® Plus due to high instances of *N*-2 stutter, an artifact which lacks thorough characterization in the literature. Analysis of one- and two-contributor PowerPlex® 16 HS samples with NOCIt was less affected by the presence of *N*-2 stutter because NOCIt utilizes calibration data to model baseline noise.

Sample complexity was also found to affect the accuracy of the number of contributors estimate, where a decrease in the accuracy of all methods tested was observed as the mass of the minor contributor decreased, and at extreme minor contributor masses (i.e., 0.01 ng), the accuracy of all methods examined did not exceed 40%. The accuracy of NOCIt was found to be 100% when the

minor component contributed the equivalent of approximately 10 cells' worth of DNA. Cumulatively, these results serve as an indication that forensic DNA samples containing low-target quantities may need to be interpreted using multiple assumptions on the number of contributors, because the assumption on the number of contributors is known to affect the conclusion in certain casework scenarios. Further studies to elucidate the impact of the assumption on the number of contributors are warranted.

References

1. Gill P, Jeffreys AJ, Werrett DJ. Forensic application of DNA "fingerprints". Nature 1985;318(6046):577–9.
2. Butler JM. Short tandem repeat typing technologies used in human identity testing. Biotechniques 2007;43(4 Suppl):Sii–Sv.
3. Budowle B, Eisenberg AJ, van Daal A. Low copy number typing has yet to achieve "general acceptance". Forensic Sci Int Genet Suppl Ser 2009;2(1):551–2.
4. Buckleton J, Gill P. Further Comment on "Low copy number typing has yet to achieve "general acceptance"" by Budowle, B., et al., 2009. Forensic Sci Int Genet Suppl Ser 2009;2(1):551–552. Forensic Sci Int Genet Suppl Ser 2011;5(1):7–11.
5. Budowle B, Eisenberg AJ, van Daal A. Validity of low copy number typing and applications to forensic science. Croat Med J 2009;50(3):207–17.
6. Pfeifer CM, Klein-Unseld R, Klintschar M, Wiegand P. Comparison of different interpretation strategies for low template DNA mixtures. Forensic Sci Int Genet 2012;6(6):716–22.
7. Butler JM. Low-level DNA and complex mixtures. Advanced topics in forensic DNA typing: interpretation. San Diego, CA: Academic Press, 2015;159–82.
8. Haned H. *Forensim*: an open-source initiative for the evaluation of statistical methods in forensic genetics. Forensic Sci Int Genet 2011;5(4):265–8.
9. Balding DJ. Evaluation of mixed-source, low-template DNA profiles in forensic science. Proc Natl Acad Sci 2013;110(30):12241–6.
10. Balding DJ, Buckleton J. Interpreting low template DNA profiles. Forensic Sci Int Genet 2009;4(1):1–10.
11. Steele CD, Balding DJ. Statistical evaluation of forensic DNA profile evidence. Annu Rev Stat Appl 2014;1(1):361–84.
12. Taylor D, Bright J-A, Buckleton J. The interpretation of single source and mixed DNA profiles. Forensic Sci Int Genet 2013;7(5):516–28.
13. Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, et al. Validating TrueAllele® DNA mixture interpretation. J Forensic Sci 2011;56(6):1430–47.
14. Cowell RG, Graversen T, Lauritzen SL, Mortera J. Analysis of forensic DNA mixtures with artefacts. J R Stat Soc Ser C Appl Stat 2015;64(1):1–48.
15. Bleka Ø, Storvik G, Gill P. *EuroForMix*: an open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. Forensic Sci Int Genet 2016;21:35–44.
16. Benschop CCG, Haned H, Jeurissen L, Gill PD, Sijen T. The effect of varying the number of contributors on likelihood ratios for complex DNA mixtures. Forensic Sci Int Genet 2015;19:92–9.
17. Paoletti DR, Doom TE, Krane CM, Raymer ML, Krane DE. Empirical analysis of the STR profiles resulting from conceptual mixtures. J Forensic Sci 2005;50(6):1361–6.
18. Buckleton JS, Curran JM, Gill P. Towards understanding the effect of uncertainty in the number of contributors to DNA stains. Forensic Sci Int Genet 2007;1(1):20–8.
19. Coble MD, Bright J-A, Buckleton JS, Curran JM. Uncertainty in the number of contributors in the proposed new CODIS set. Forensic Sci Int Genet 2015;19:207–11.
20. Haned H, Pène L, Lobry JR, Dufour AB, Pontier D. Estimating the number of contributors to forensic DNA mixtures: does maximum likelihood perform better than maximum allele count? J Forensic Sci 2011;56(1):23–8.
21. Perez J, Mitchell AA, Ducasse N, Tamariz J, Caragine T. Estimating the number of contributors to two-, three-, and four-person mixtures containing DNA in high template and low template amounts. Croat Med J 2011;52(3):314–26.
22. Biedermann A, Bozza S, Konis K, Taroni F. Inference about the number of contributors to a DNA mixture: comparative analyses of a Bayesian network approach and the maximum allele count method. Forensic Sci Int Genet 2012;6(6):689–96.
23. Tvedebrink T. On the exact distribution of the numbers of alleles in DNA mixtures. Int J Legal Med 2013;128(3):427–37.
24. Swaminathan H, Grgicak CM, Medard M, Lun DS. NOCIt: a computational method to infer the number of contributors to DNA samples analyzed by STR genotyping. Forensic Sci Int Genet 2015;16:172–80.
25. Applied Biosystems. Quantifiler® Duo DNA quantification kit user's manual. Foster City, CA: Applied Biosystems, 2012.
26. Grgicak CM, Urban ZM, Cotton RW. Investigation of reproducibility and error associated with qPCR methods using Quantifiler® Duo DNA quantification kit. J Forensic Sci 2010;55(5):1331–9.
27. Life Technologies. AmpFlSTR® Identifiler® Plus PCR amplification kit user guide. Carlsbad, CA: Life Technologies, 2014.
28. Promega Corporation. PowerPlex® 16 HS system technical manual. Madison, WI: Promega Corporation, 2014.
29. Swaminathan H, Lun D, Medard M, Norsworthy S, Rowan K, Wellner G, et al. NOCIt: a method for determining the number of contributors. Technical Manual v1.5.1, 2014; http://sites.bu.edu/grgicak/software/ (accessed October 24, 2016).
30. http://www.cstl.nist.gov/strbase/training/ISHI2012-MixtureWorkshop-Profile1.pdf (accessed February 2, 2016).
31. Mönich UJ, Duffy K, Médard M, Cadambe V, Alfonse LE, Grgicak C. Probabilistic characterisation of baseline noise in STR profiles. Forensic Sci Int Genet 2015;19:107–22.
32. Haned H. A tutorial for the R package *forensim*, 2013; http://forensim.r-forge.r-project.org/misc/forensim-tutorial.pdf (accessed September 24, 2015).
33. Benschop CCG, Haned H, de Blaeij TJP, Meulenbroek AJ, Sijen T. Assessment of mock cases involving complex low template DNA mixtures: a descriptive study. Forensic Sci Int Genet 2012;6(6):697–707.

Additional information and reprint requests:
Catherine M. Grgicak, Ph.D.
Boston University School of Medicine
Biomedical Forensic Sciences
72 East Concord Street, Room R806B
Boston, MA 02118
E-mail: cgrgicak@bu.edu

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 List of test samples evaluated in each of the three phases of this study.