Forensic Science International: Genetics 34 (2018) 11–24

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Internal validation of STRmix™ – A multi laboratory response to PCAST

Jo-Anne Bright[a,*], Rebecca Richards[a], Maarten Kruijver[a], Hannah Kelly[a], Catherine McGovern[a], Alan Magee[b], Andrew McWhorter[c], Anne Ciecko[d], Brian Peck[e], Chase Baumgartner[f], Christina Buettner[g], Scott McWilliams[g], Claire McKenna[h], Colin Gallacher[i], Ben Mallinder[i], Darren Wright[j], Deven Johnson[k], Dorothy Catella[l], Eugene Lien[m], Craig O'Connor[m], George Duncan[n], Jason Bundy[o], Jillian Echard[p], John Lowe[q], Joshua Stewart[r], Kathleen Corrado[s], Sheila Gentile[s], Marla Kaplan[t], Michelle Hassler[u], Naomi McDonald[v], Paul Hulme[w], Rachel H. Oefelein[x], Shawn Montpetit[y], Melissa Strong[y], Sarah Noël[z], Simon Malsom[A], Steven Myers[B], Susan Welti[C], Tamyra Moretti[D], Teresa McMahon[E], Thomas Grill[F], Tim Kalafut[G], MaryMargaret Greer-Ritzheimer[H], Vickie Beamer[I], Duncan A. Taylor[J,K], John S. Buckleton[a,L]

[a] *Institute of Environmental Science and Research Limited, Private Bag 92021, Auckland, 1142, New Zealand*
[b] *Forensic Science Ireland, Ireland*
[c] *Texas Department of Public Safety, Houston Laboratory, United States*
[d] *Midwest Regional Forensic Laboratory, Andover, MN, United States*
[e] *Centre of Forensic Sciences, Toronto, Canada*
[f] *Texas Department of Public Safety, Austin Laboratory, United States*
[g] *Wyoming State Crime Laboratory, United States*
[h] *Austin Police Department, City of Austin, TX, United States*
[i] *Scottish Police Authority (SPA), United Kingdom*
[j] *Idaho State Police Forensic Services, United States*
[k] *Sacramento District Attorney's Office Laboratory of Forensic Services, CA, United States*
[l] *Oakland County Sheriff's Office, MI, United States*
[m] *New York City Office of Chief Medical Examiner (OCME), United States*
[n] *Broward Sheriff's Office Crime Laboratory, FL, United States*
[o] *Florida Department of Law Enforcement, United States*
[p] *Connecticut DESPP Division of Scientific Services, United States*
[q] *Key Forensic Services Ltd., Warrington Laboratory, United Kingdom*
[r] *Texas Department of Public Safety, Corpus Christi Laboratory, United States*
[s] *Onondaga County Center for Forensic Sciences, NY, United States*
[t] *Oregon State Police Laboratory (OSP), United States*
[u] *San Diego County Sheriff's Regional Crime laboratory, United States*
[v] *Texas Department of Public Safety, Lubbock Laboratory, United States*
[w] *Cellmark Forensic Services, United Kingdom*
[x] *DNA Labs International, United States*
[y] *San Diego Police Department Crime Laboratory, CA, United States*
[z] *Laboratoire de sciences judiciaires et de médecine légale (LSJML) Montréal, Canada*
[A] *Key Forensic Services Ltd., Norwich Laboratory, United Kingdom*
[B] *California Department of Justice Bureau of Forensic Services, United States*
[C] *Department of Forensic Sciences Laboratory, Washington DC (DFS), United States*
[D] *Federal Bureau of Investigation (FBI), United States*
[E] *Forensic Science Northern Ireland, Northern Ireland*
[F] *Erie County Central Services Laboratory, Buffalo, NY, United States*
[G] *US Army Criminal Investigation Laboratory (USACIL), United States*
[H] *DuPage County Sheriff's Crime Laboratory, IL, United States*
[I] *Scottsdale Police Department Crime Laboratory, AZ, United States*
[J] *Forensic Science South Australia, 21 Divett Place, Adelaide, SA 5000, Australia*
[K] *School of Biological Sciences, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia*
[L] *University of Washington, Department of Biostatistics, Seattle, WA 98195, United States*

* Corresponding author.
E-mail address: jo.bright@esr.cri.nz (J.-A. Bright).

https://doi.org/10.1016/j.fsigen.2018.01.003
Received 20 August 2017; Received in revised form 29 November 2017; Accepted 6 January 2018
Available online 08 January 2018
1872-4973/

ARTICLE INFO

Keywords:
PCAST
STRmix
Forensic DNA
Probabilistic genotyping
Continuous models
Validation

ABSTRACT

We report a large compilation of the internal validations of the probabilistic genotyping software STRmix™. Thirty one laboratories contributed data resulting in 2825 mixtures comprising three to six donors and a wide range of multiplex, equipment, mixture proportions and templates. Previously reported trends in the LR were confirmed including less discriminatory LRs occurring both for donors and non-donors at low template (for the donor in question) and at high contributor number. We were unable to isolate an effect of allelic sharing. Any apparent effect appears to be largely confounded with increased contributor number.

1. Introduction

In 2016, the President's Council of Advisors on Science and Technology (PCAST) issued a report [1] and subsequently an addendum [2]. This report discussed a number of forensic disciplines. Included amongst these was the interpretation of complex DNA mixtures. PCAST defined a complex mixture as any profile with three or more donors. The report noted perceived limits to the proof of validity of the use of probabilistic genotyping (PG) in some situations as of September 2016. In particular they highlighted gaps regarding high ratio and high contributor number mixtures. PCAST considered validity proven for mixtures containing "*three contributors where the person of interest comprises at least 20% of the sample*." [2]. They noted that the "*few studies that have explored 4- or 5-person mixtures often involve mixtures that are derived from only a few sets of people (in some cases, only one)*." [2]. They call for the expansion of empirical studies, testing the validity and reliability of PG methods across a broader relevant range of profile types.

PCAST limited themselves for proof of validity to empirical studies published in the peer reviewed literature. There are a number of published reports describing the validation of various probabilistic genotyping software by the developers. These include the New York City Office of Chief Medical Examiner's FST Tool [3], TrueAllele® [4], and STRmix™ [5]. More recently the validation of GenoProof Mixture 3 [6] and *Kongoh* [7] has been reported.

PCAST also perceived there was a gap in "*the need for clarity about the scientific standards for the validity and reliability of forensic methods*." [1]. The Scientific Working Group on DNA Analysis Methods (SWGDAM) [8] and International Society for Forensic Genetics (ISFG) [9] have both published comprehensive guidelines that inform how to test a probabilistic genotyping system to ensure reliability and validity of results.

At the time of the PCAST report there was a considerable number of empirical studies already undertaken by various laboratories who had implemented, or were in the process of implementing, STRmix™. These followed the SWGDAM guidelines [10,11]. They were not published in the peer reviewed literature largely because it is the policy of many journals not to publish such material. Some of these studies are already in the public domain on websites (see for example [12,13]).

Since the appearance of the PCAST report, the Federal Bureau of Investigation Laboratory, Quantico, has published its STRmix™ internal validation in the peer reviewed literature [14], also in accordance with the SWGDAM guidelines. This publication reports 277 mixtures with two to five donors and a range of mixture ratios and templates.

In this work we report a further study of 2825 mixtures compiled from 31 laboratories (including multi laboratory systems) who are using STRmix™ in casework (28/31) or currently validating STRmix™ for future use in casework (3/31). Mixtures of three, four, five, and six contributors were specifically targeted in order to address the criticisms of PCAST.

We aim to specifically address the deficiencies described by PCAST in their report by addressing the following points:

(1) How well does the method perform as a function of the number of contributors to the mixture? How well does it perform when the number of contributors to the mixture is *unknown*?

(2) How does the method perform as a function of the number of alleles shared among individuals in the mixture? Relatedly, how does it perform when the mixtures include related individuals?

(3) How well does the method perform – and how does accuracy degrade – as a function of the absolute and relative amounts of DNA from the various contributors?

We address point 1 in experiment 1 by analysing all submitted mixtures assuming the *apparent* number of contributors. The apparent number of contributors (N) was determined blind by the submitting laboratory following their own standard operating procedures. Note that this resulted in all six person mixtures being analysed as assuming less than six. Additionally, we have assumed N + 1 for a subset of the data within experiment 2. Point 2 we address by interrogating the data in experiment 1 with respect to the amount of allelic sharing. Point 3 we address by conducting H_p and H_d true tests on mixtures in experiment 1.

In this work the developers of STRmix™ did not generate or choose the data that was analysed by individual (non-developing) laboratories and they have not censored any data from the results. This adheres to the call by PCAST for work to be carried out in conjunction between developers and non-developing organisations.

There is a fourth point to the list in the PCAST report:

(4) Under what circumstances – and why – does the method produce results (random inclusion probabilities) that differ substantially from those produced by other methods?

We do not address point (4) within this paper, however work is ongoing to address it across a number of continuous and semi-continuous platforms.

2. Methods

2.1. Data submission

Participating laboratories submitted ground truth known profiles originating from three to six contributors that had previously been interpreted as part of their STRmix™ internal validation studies. Profiles were submitted as analysed data in the form of text or Excel files. In addition, laboratories provided reference profiles for the known contributors, their validated laboratory specific settings, and the apparent number of contributors to each profile. The apparent number of contributors was determined by the submitting laboratories following their own standard operating procedures. The apparent number of contributors was used as the true number of contributors to a crime profile is never known.

2.1.1. Data description

Apparent three, four and five person mixtures were interpreted by staff at ESR (New Zealand) using STRmix™ V2.5.02. No apparent single source or two person mixtures were interpreted as PCAST, perhaps erroneously, decreed foundational validity to be already established for these [1]. In total there were 2825 mixtures interpreted from 31 different laboratories generated using eight different STR multiplexes and analysed on two different types of capillary electrophoresis (CE) instruments.

The STRmix™ settings used for the interpretation were those determined by the contributing laboratory. These included per allele stutter ratios (back and forward, where determined), allele and stutter peak height variance distributions, analytical thresholds, saturation,

and drop-in parameters. For each interpretation, eight MCMC chains of 100,000 burn-in accepts and 50,000 post burn-in accepts were used.

The number of profiles submitted, multiplex, PCR cycle number, CE instrument used, and number of mixtures interpreted for each participating laboratory are provided in Table 1. Note some laboratories submitted profiles generated using more than one multiplex (kit) and some were multi laboratory systems, submitting profiles from different laboratories within the one system. Many of the laboratories undertook dilution series to prepare mixtures for interpretation. These were typically made by taking DNA from a few donors, often staff members, and mixing them in different combinations and ratios. PCAST noted that "In human molecular genetics, an experimental validation of an important diagnostic would typically involve hundreds of distinct samples." (PCAST pg 81). Each different combination of genotypes is a unique contributor combination.

The number of the unique contributor combinations for each mixture type is given in Table 1. For example, there were twelve combinations of different contributors for the apparent three person mixtures submitted by Lab 01. In total there were 25 apparent three person mixtures from Lab 01, hence 12/25 in Table 1. For all laboratories, there were 205 unique three contributor profiles, 132 unique four contributor profiles, and 14 unique five contributor profiles. Within the STRmix™ deconvolution, template is modelled per contributor [11]. The mode of the post burn-in proposals for template per contributor was used to calculate mixture proportion. The mixture proportions as determined by STRmix™ (sorted by ascending proportion for contributor 1, constrained as the 'major' contributor) are plotted for each apparent N in Fig. 1. At least one contributor in 69.5% of the apparent three person mixtures, 96.5% of the apparent four person mixtures and all of the apparent five person mixtures contained less than 20% of the sample.

PCAST calls for an investigation to be conducted into how a method "*performs as a function of the number of alleles shared among individuals in the mixture*". In Fig. 2 we provide the distribution of allele sharing for known contributors in the mixtures, broken down by the true number of contributors to a mixture. Allele sharing (AS) is defined as the fraction of alleles in the donors collectively that appear in two or more donor genotypes. The upper tail (> 0.80 proportion AS) for the three and four contributor mixtures are a known family group consisting of a mother, father, and their two biological children that was investigated by one participating laboratory.

2.2. Experiment 1

For each profile, likelihood ratios (*LR*s) were calculated for the true donors and 10,000 false donors. The profiles of the 10,000 non-donors were created by simulation using the FBI Caucasian allele frequencies for each multiplex. All *LR*s were calculated using the Caucasian allele

Table 1
A list of the contributing laboratories, multiplex (kit) used, PCR cycle number, and CE instrument. The total number of mixtures interpreted per laboratory are sorted by apparent number of contributors with the number of unique contributor combinations and minimum minor proportion as determined by STRmix™ indicated.

					Number of each mixture type Unique contributor combinations/total (Minimum minor contribution)		
Lab	Samples submitted (true N)	Kit	Cycle Number	CE	Apparent 3p	Apparent 4p	Apparent 5p
L01	$N_3 = 24, N_4 = 23$	Fusion 5C	28	3130	12/25(7%)	12/22(7%)	–
L02	$N_3 = 19, N_4 = 24$	Identifiler™ Plus	28	3500	4/21(6%)	3/22(6%)	–
L03	$N_3 = 88, N_4 = 128, N_5 = 48$	GlobalFiler™	29	3500	5/87(3%)	6/161(< 1%)	2/16(5%)
L04	$N_3 = 3, N_4 = 3$	NGM SElect™	30	3130	1/3(10%)	1/3(6%)	–
L05	$N_3 = 39, N_4 = 37$	Fusion 6C	29	3130	5/50(3%)	4/26(< 1%)	–
L06	$N_3 = 28, N_4 = 69$	Identifiler™ Plus	28	3130	4/67(28%)	2/30(12%)	–
L07	$N_3 = 29, N_4 = 30$	Identifiler™ Plus	28	3130	4/36(2%)	1/23(2%)	–
L08	$N_3 = 19, N_4 = 20$	Fusion 6C	29	3500	2/24(7%)	1/15(4%)	–
L09	$N_3 = 28, N_4 = 8, N_5 = 6$	Fusion 5C	30	3500	4/28 (1%)	2/8(2%)	1/6(6%)
	$N_3 = 22, N_4 = 22$	Identifiler™ Plus	29	3500	1/22 (1%)	1/22 (2%)	–
L10	$N_3 = 29, N_4 = 52, N_5 = 12$	GlobalFiler™	28	3500	4/64 (3%)	4/29 (1%)	–
L11	$N_3 = 69, N_4 = 42$	GlobalFiler™	28	3500	2/69 (< 1%)	2/42 (1%)	–
L12	$N_3 = 28, N_4 = 32$	NGM SElect™	29	3500	2/38 (5%)	1/22 (5%)	–
L13	$N_3 = 3, N_4 = 3$	NGM SElect™	30	3130	1/3 (9%)	1/3 (3%)	–
	$N_3 = 3, N_4 = 3$	PowerPlex® ESI17 Pro	30	3130	1/3 (13%)	1/3 (6%)	–
L14	$N_3 = 10, N_4 = 13$	PowerPlex® 16 HS	30	3130	2/16 (7%)	1/7 (5%)	–
L15	$N_3 = 26$	PowerPlex® ESI17 Fast	30	3130	11/26 (2%)	–	–
	$N_3 = 28$	PowerPlex® ESI17 Fast	30	3500	11/28 (2%)	–	–
L16	$N_3 = 29, N_4 = 11$	Identifiler™ Plus	28	3130	9/38 (4%)	1/2 (5%)	–
L17	$N_3 = 26, N_4 = 32$	GlobalFiler™	29	3500	2/32 (4%)	1/26 (1%)	–
L18	$N_3 = 97, N_4 = 46$	Fusion 5C	29	3130	7/108 (7%)	3/35 (2%)	–
L19	$N_3 = 28, N_4 = 30$	Identifiler™ Plus	29	3130	9/37 (3%)	15/21 (2%)	–
L20	$N_3 = 22, N_4 = 23, N_5 = 12$	GlobalFiler™	29	3500	9/42 (< 1%)	4/13 (5%)	1/2 (1%)
L21	$N_3 = 43, N_4 = 39$	Fusion 6C	29	3500	14/59 (4%)	9/23 (1%)	–
L22	$N_3 = 62, N_4 = 65, N_5 = 11$	GlobalFiler™	29	3500	27/69 (3%)	25/64 (1%)	2/5 (7%)
L23	$N_3 = 72, N_4 = 64$	Fusion 6C	29	3500	6/83 (1%)	4/53 (< 1%)	–
	$N_3 = 159, N_4 = 60$	Identifiler™ Plus	28	3130	4/161 (1%)	3/58 (< 1%)	–
L24	$N_3 = 35, N_4 = 36$	GlobalFiler™	29	3500	4/37 (3%)	3/34 (2%)	–
L25	$N_3 = 20, N_4 = 24$	GlobalFiler™	29	3500	1/20 (5%)	1/24 (6%)	–
L26	$N_3 = 18, N_4 = 12$	Identifiler™ Plus	28	3130	17/25 (6%)	3/5 (< 1%)	–
L27	$N_3 = 51, N_4 = 42$	Identifiler™ Plus	28	3500	5/71 (3%)	2/22 (< 1%)	–
L28	$N_3 = 12, N_4 = 77, N_5 = 76, N_6 = 65$	Fusion 5C	29	3500	6/24 (3%)	7/151 (< 1%)	6/55 (< 1%)
L29	$N_3 = 52, N_4 = 52$	GlobalFiler™	28	3500	2/53 (3%)	1/51 (1%)	–
L30	$N_3 = 31, N_4 = 42$	GlobalFiler™	29	3500	4/42 (4%)	3/31 (< 1%)	
L31	$N_3 = 63, N_4 = 99, N_5 = 17$	GlobalFiler™	29	3500	3/80 (1%)	4/85 (< 1%)	2/14 (< 1%)
		TOTAL Number of each mixture type unique combinations/total (minimum minor contribution)			205/1591 (< 1%)	132/1136 (< 1%)	14/98 (< 1%)

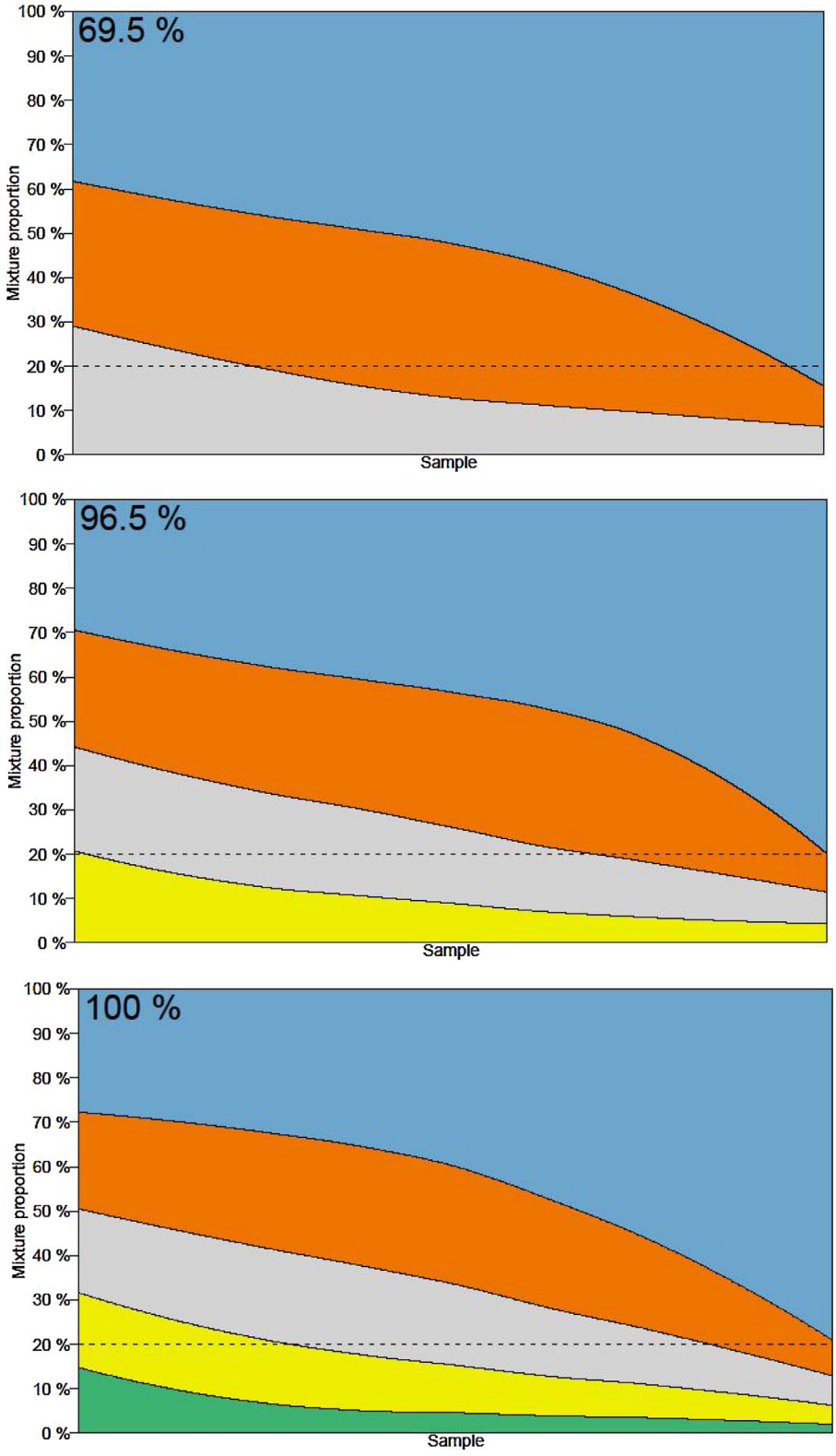

Fig. 1. Mixture proportions as calculated by STRmix™ and sorted by ascending proportion plotted by apparent N where 1a is apparent three, 1b apparent four and 1c apparent five N. Plots are smoothed for improved readability.

frequencies from the FBI expanded CODIS core set [15] and a theta (F_{ST}) of 0.01. The propositions considered were:

H_p: the DNA originated from the person of interest (either true or false donor) and N-1 unknown contributors

H_d: the DNA originated from N unknown contributors

where N was the apparent number of contributors.

Average peak height (*APH*) was calculated for each contributor by averaging the peak heights of the unmasked alleles (not shared between contributors and not in back stutter positions of any other contributor alleles). Alleles that had dropped out were assigned a height of half the laboratory's analytical threshold (AT).

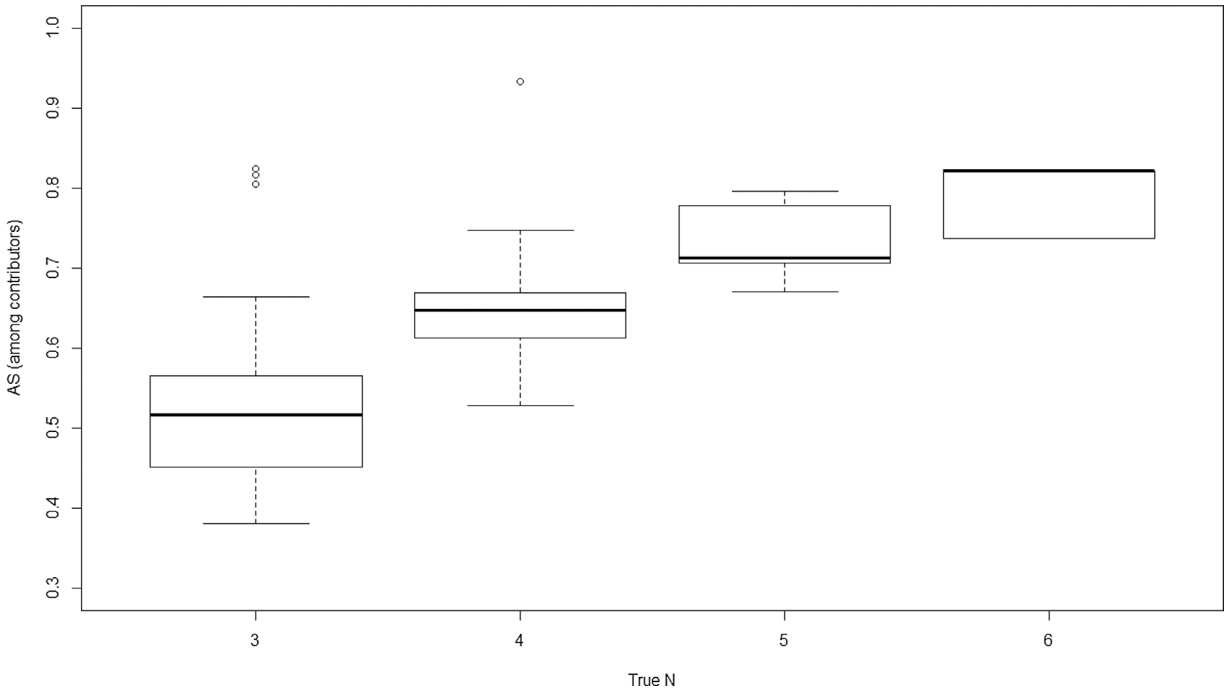

Fig. 2. Distribution of allele sharing (AS) for known contributors to mixtures, plotted by true N.

2.3. Experiment 2

For one laboratory the three and four contributor profiles were analysed at both the apparent number of contributors (N) and one greater (N + 1). For these mixtures, apparent N was the same as known N. In practise, when analysed as N + 1 a non-existent contributor with true mixture proportion 0 has been added to reflect this ambiguous contributor being present at trace amounts. The mixture proportion for this additional contributor was constrained to be low, but not necessarily zero, using the informed mixture proportion prior function in STRmix™ [16]. The *LR*s for the true donors and 10,000 non-donors were assigned as per Experiment 1.

3. Results

3.1. Data review

The summary statistics for each interpretation were reviewed prior to review of the *LR*. These statistics included the Gelman-Rubin convergence statistic, average $\log_{10}$(likelihood) of the post burn-in MCMC, the average of the post burn-in allele variance parameter, and the average of the post burn-in stutter variance parameter. These values can be used as diagnostics of the interpretation, to check for adequate MCMC convergence. They are designed to help assess a STRmix™ deconvolution result. No profiles required reinterpretation based on the review of the diagnostics.

The *LR*s were also reviewed as part of data quality checks. Large inclusionary *LR*s ($LR > > 1$) for false contributors and exclusionary *LR*s ($LR < 1$) for true contributors where the *APH* was relatively high were investigated. For any given mixture, there is a chance that a given false contributor will have sufficient matching alleles, by chance, to give an $LR > 1$. Likelihood ratios for false contributors above 10,000 are provided in Table 2. Following Taylor et al. [17];

1) The average *LR* for false contributors should be about 1.
2) The probability of observing a likelihood ratio of *x* or larger from an unrelated non-donor is no more than 1 in *x*.

These two statements fo rm the basis for assessing false contributor tests. In an experiment on 10,000 false contributors we would expect approximately one $LR \geq 10{,}000$, plausibly 10 above 1000 and 100 above 100. This work reports the comparison of approximately 20 million false contributors. The average *LR* for all false contributors is approximately 0.12. The reason that this average is below one is because the genotypes that would lead to the highest *LR*s (and so contribute significantly to the average) were not happened across in the number of H_d true tests performed.

The fraction of allele sharing for the twenty highest false contributors ranged from 0.61 to up to 0.98 of the alleles with the mixture (Table 2).

False exclusions were observed for known contributors where the apparent number of contributors was fewer than the ground truth

Table 2
Summary of large inclusionary *LR*s for false contributors and percentage of overlapping alleles.

Number	Kit	Apparent N	Known N	*LR*	Fraction of allele sharing
1	GlobalFiler™	3	3	505,924	0.81
2	Identifiler Plus™	3	3	379,716	0.90
3	GlobalFiler™	4	4	197,907	0.98
4	GlobalFiler™	3	4	134,486	0.83
5	GlobalFiler™	4	4	88,022	0.98
6	GlobalFiler™	4	5	53,019	0.93
7	Fusion 6C	3	3	47,062	0.85
8	Fusion 5C	3	3	43,065	0.78
9	Fusion 5C	3	3	26,874	0.80
10	GlobalFiler™	3	3	19,340	0.67
11	Fusion 5C	3	3	17,582	0.61
12	Identifiler Plus™	3	4	16,995	0.80
13	Fusion 5C	4	4	15,765	0.80
14	Identifiler Plus™	3	3	14,446	0.87
15	NGM SElect™	3	4	13,717	0.78
16	GlobalFiler™	4	5	12,135	0.93
17	Fusion 5C	4	6	11,188	0.93
18	Fusion 5C	3	3	10,896	0.80
19	Fusion 5C	3	3	10,309	0.82
20	Identifiler Plus™	3	3	10,298	0.80

number of contributors. This was an expected result [18,19]. By way of explanation we present an example of a true five contributor mixture interpreted assuming four contributors. Fig. 3 is a stylised electropherogram for one locus (SE33) with peaks and their corresponding height. STRmix™ has modelled the minor peaks as stutters of the eight alleles all above 800 rfu. Assuming four contributors and eight alleles, each contributor must be heterozygous at this locus. One known contributor who is homozygous at this locus (genotype 18,18) is therefore excluded ($LR_{SE33} = 0$) as a contributor under the assumption of four contributors. A second individual (genotype 12,23.2) is a poor fit to the profile assuming four contributors given the large peak imbalance for these alleles resulting in a low weight and subsequent *LR* at this locus ($LR_{SE33} = 0.01$).

False exclusions were also observed due to human error if, for example, an incorrect reference profile was supplied. Human errors were all corrected and the *LR*s reassigned. Another common reason for a false exclusion was due to the lack of separation of alleles during capillary electrophoresis. This occurred when peaks that differed by one base pair (for example a 9.3/10 at TH01) were not separated sufficiently during electrophoresis and one was subsequently not designated at analysis [14]. In all identified occasions an allele corresponding with a minor contributor was 'hidden' within the shoulder of an allele from a major contributor. Affected loci were identified by reviewing the electropherogram, and the locus was subsequently ignored during the interpretation.

3.2. Results for experiment 1

Violin plots [20] showing the densities of $\log_{10}(LR)$ per *APH* range are provided in Fig. 4 through 6 for apparent three, four and five contributor mixtures, respectively. The percentage of non-contributors giving $LR = 0$ is given at the bottom of each plot. The plots show the general trends for both H_p and H_d results.

Plots of $\log_{10}(LR)$ versus *APH* for all mixtures are given in the Supplementary material Figs. S1 through S9, plotted by apparent number of contributors. These plots are also separated into H_p true (*LR*s for true donors) and H_d true results (*LR*s for 10,000 false donors) and H_p and H_d true combined in order to help visualise the trends. In order to facilitate comparison between plots the axis scales have been retained for the same N. For the H_p true results where apparent N differed from the true N these results are indicated with a different plotting symbol. *LR* results of 0 (exclusions) have been plotted at -40 on the $\log_{10}$ scale. Normalisation of the CE platform (3130 versus 3500) had no effect on the trends present in the data and is not shown.

The vertical line of points in Fig. S8 at 50 rfu where $\log_{10}(LR) > 1$ are two siblings from a family study that included their biological father and mother. Due the complete allele sharing with both parents the *APH* for both siblings were calculated at half the AT, which is artificially low.

Fig. 4, Fig. 5 and Fig. 6 show the same trends as seen in previous work [14,21], with the addition of information regarding the consequence of over or underestimating the number of contributors. With increased information present within the profile (either by greater amounts of DNA, or by fewer contributors) the power to discriminate contributors from non-contributors increases, and there is a divergence of the *LR* from neutrality. Also consistent with previous findings [18], the underestimation of the number of contributors tends to either have little effect on the *LR* or will tend to exclude known contributors. This occurs because genotype sets possessing unreal allele pairings are forced to be given weight within the analysis. Interestingly this exclusionary effect was reduced as mixture complexity increased to the point that there were no exclusions produced from underestimating the number of contributors in five person mixtures (Fig. S1). We surmise that this is an effect of the increased allele sharing generally seen in higher order mixtures (Fig. 2) meaning that there are increased opportunities for genotype sets to possess the genotypes of the known contributors, even when their number is underestimated.

A plot of $\log_{10}(LR)$s for profiles generated using Identifiler™ Plus 28 cycles analysed on a 3130 or 3500 are plotted in Figs. S10 and S11 for the apparent three and four person mixtures, respectively (Supplementary material). As a visual aid we have added smoothed trend lines (LOWESS lines) for instrument type. These trend lines give a rough idea of the relationship between $\log_{10}(LR)$ and *APH* for different cases. Any trend line is a compromise between smoothness and error. We did not get materially different results when trying other trend lines

Peak	Height
12	892
14	116
15	1104
17	155
18	1899
22.2	186
23.2	2334
24.2	147
25.2	1386
26.2	1508
27.2	1410
30.2	89
31.2	953

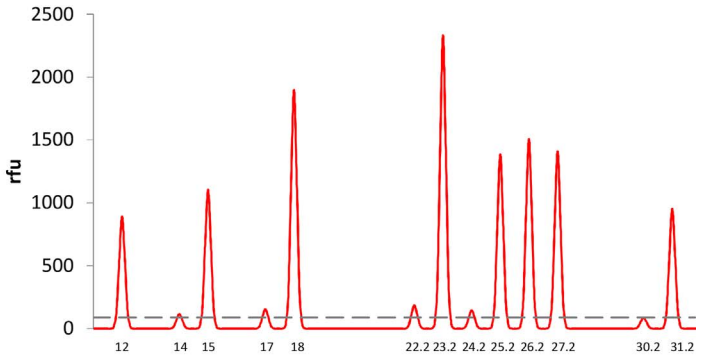

Fig. 3. Stylised locus electropherogram with tabulated peak designations and their corresponding heights for a true five person mixture interpreted assuming four contributors.

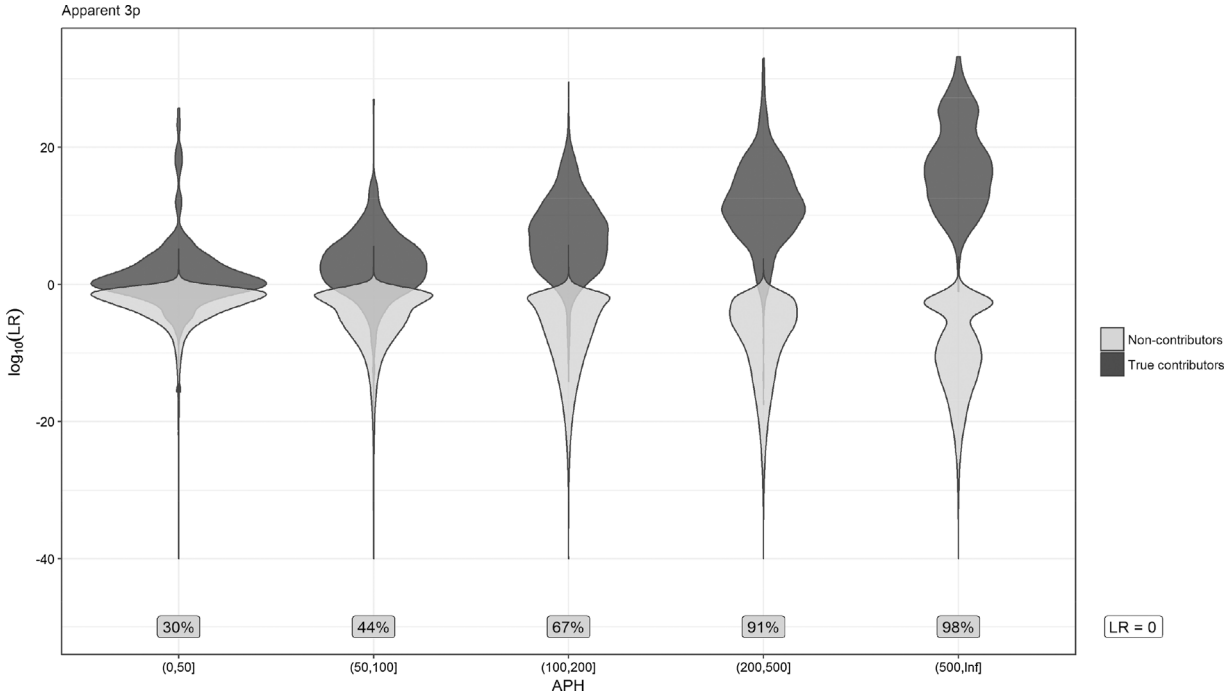

Fig. 4. Violin plot of $\log_{10}(LR)$ versus *APH* for apparent three contributor mixtures.

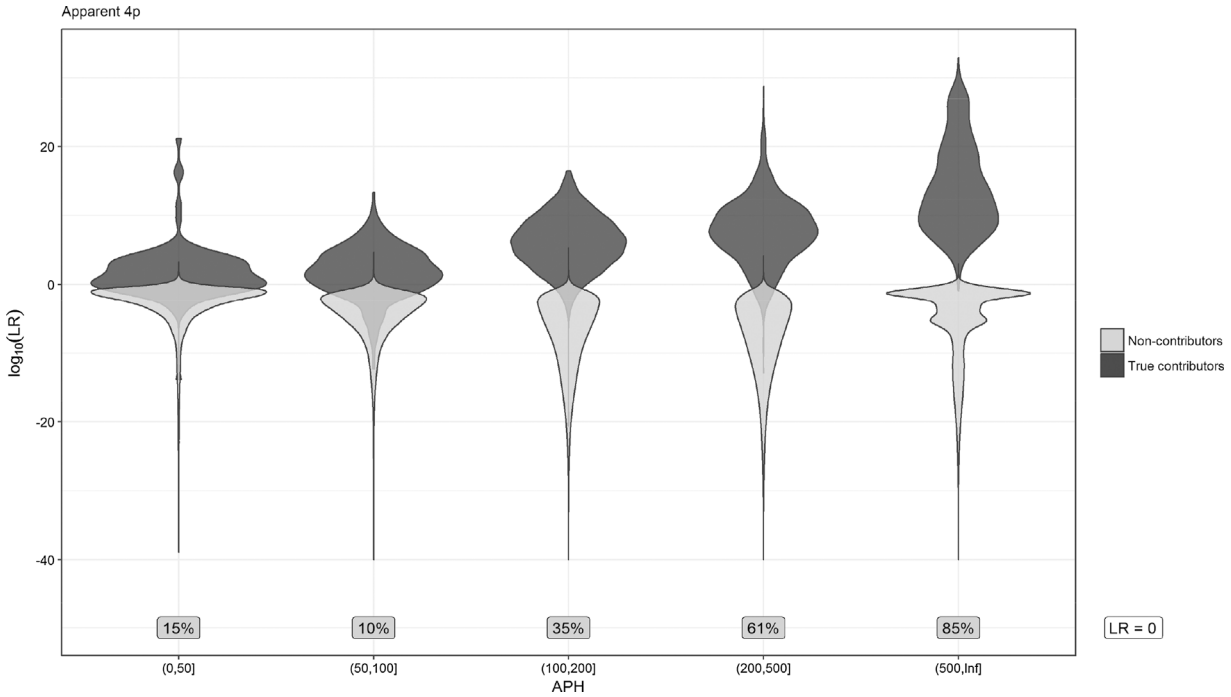

Fig. 5. Violin plot of $\log_{10}(LR)$ versus *APH* for apparent four contributor mixtures.

available in the ggplot2 package [22].

Applied Biosystems report a three- to fourfold increase in rfu scale with the 3500 models over the older Applied Biosystems 3100 and 3130 instruments [23]. This is evidenced by a general shift in the trend lines for the 3500 to the right in Figs. S10 and S11. The lines converge at high *APH* where the individual contributor profiles are likely fully represented and trend to $\log_{10}(LR) = 0$ as *APH* decreases.

Plots of $\log_{10}(LR)$s for true contributors identified by kit type are given in Figs. S12 and S13 for the apparent three and four person mixtures, respectively (Supplementary material). The LOWESS trend lines for kit type are modelled. These plots indicate the performance of the difference kits over *APH* for submitted profiles. As the profiles analysed are not the same between the different kits they are not suitable for comparing performance of the different kits. However, they do give an indication of general trends. As an example, comparing the trend lines for Identifiler™ versus GlobalFiler™ mixtures, at higher per contributor *APH* the $\log_{10}(LR)$s are higher for GlobalFiler™ profiles, most likely due to the additional loci within the GlobalFiler™ kit compared with the Identifiler™ Plus kit. $\text{Log}_{10}(LR)$ values for Identifiler™ profiles are generally higher at low contributor *APH* compared to GlobalFiler™ profiles, however. This could be due to the increased variability of the GlobalFiler™ profiles, all of which were analysed on

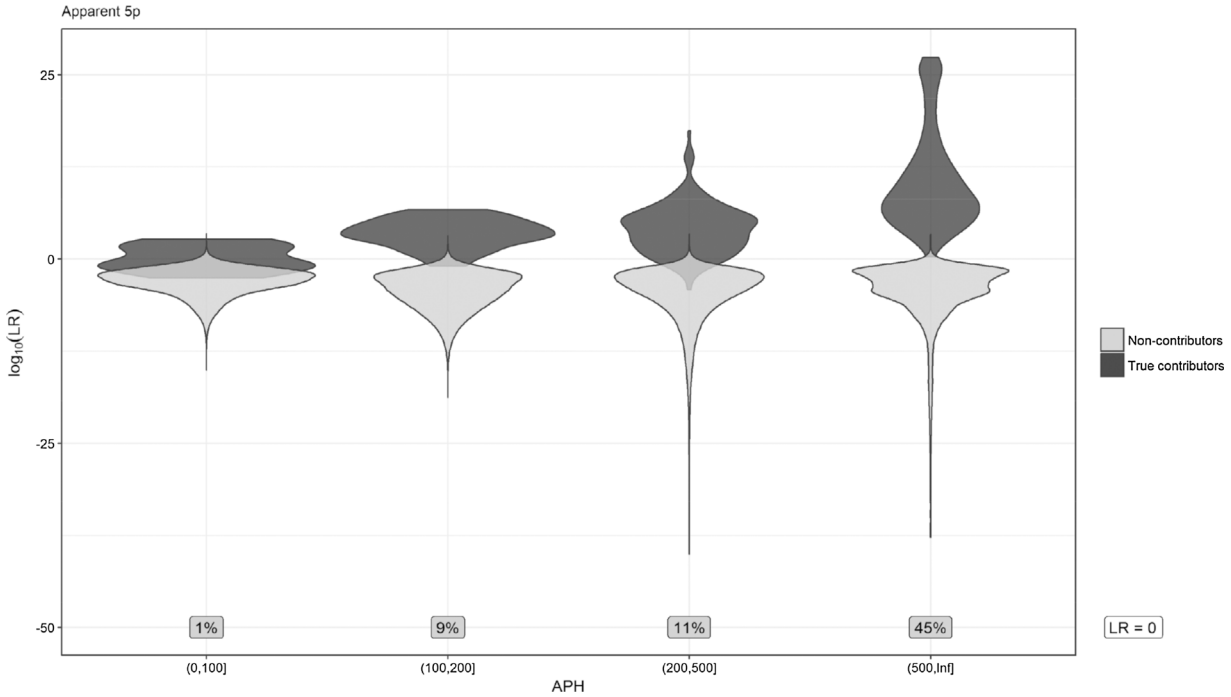

Fig. 6. Violin plot of $\log_{10}(LR)$ versus *APH* for apparent five contributor mixtures.

3500 instruments, in some cases with cycle numbers greater than 28 [24]. A comparison of the Fusion 5C and Fusion 6C trend lines illustrates the increase in discrimination achieved by adding the highly polymorphic STR locus SE33 resulting in generally higher $\log_{10}(LR)$s.

3.3. Results for experiment 2

The *LRs* for H_p true under the assumption of N and N + 1 contributors are presented in Fig. 7. Within Fig. 7 the size of the plotting symbols is relative to the contributor's proportion of the mixture. The *LRs* for H_d true are summarised in Figs. 8 and 9.

The results shown in Fig. 7 demonstrate some findings that are important for DNA mixture interpretation:

1. The general result was a decrease in the *LR* for true contributors after the assumption of an additional contributor to the mixture. The additional proposed contributor is interacting with the true contributors, diffusing the genotype weights, hence lowering the *LR*.
2. When a proposed person of interest aligns with the dominant component in a mixed DNA profile, the support for their inclusion to a mixture will not be markedly altered by an increase in the number of contributors under which the DNA profile is analysed. This is consistent with earlier findings [18].
3. Even when only donating a minor component of the total DNA, the change in *LR* produced by increasing the number of contributors is still not extreme. In no instances has an increase in the number of contributors seen an *LR* that strongly favours inclusion shift to one that favours exclusion.

We also consider the effect of contributor overestimation on H_d true tests. Fig. 8 shows the distribution of H_d true $\log_{10}(LR)$ values for three person mixtures when considered as originating from three (N) or four

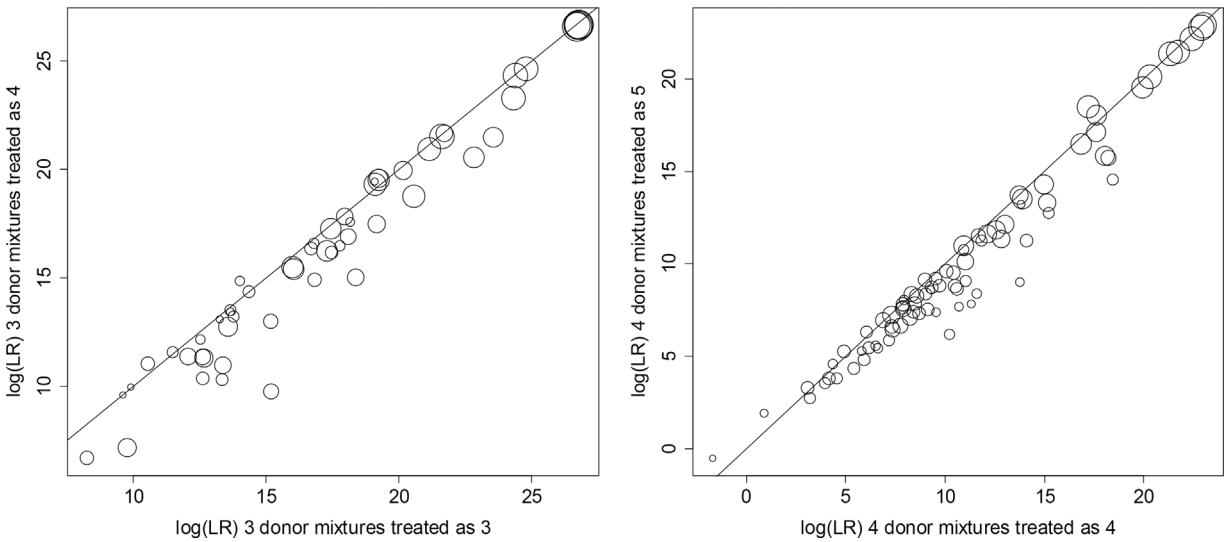

Fig. 7. The *LRs* for H_p true for three and four person mixtures from one laboratory under the assumption of N and N + 1 contributors. The $x = y$ line is shown. The size of the plotting symbol represents the mixture proportion of the donor.

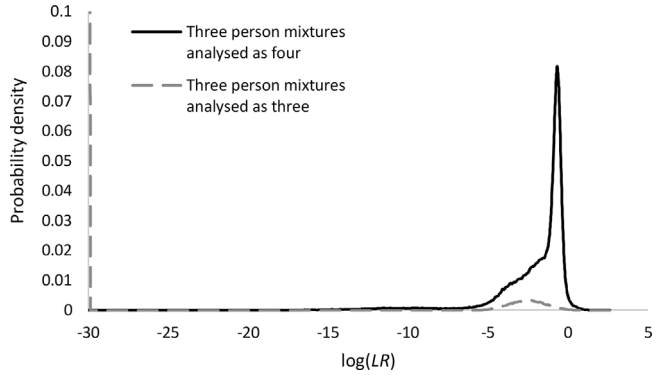

Fig. 8. The *LR*s for H_d true for three person mixtures from one laboratory under the assumption of N and N + 1. The bulk of the distribution for the three person mixtures analysed as three is at $LR = 0$ (90% of all *LR*s) represented by $\log_{10}(LR) = -30$.

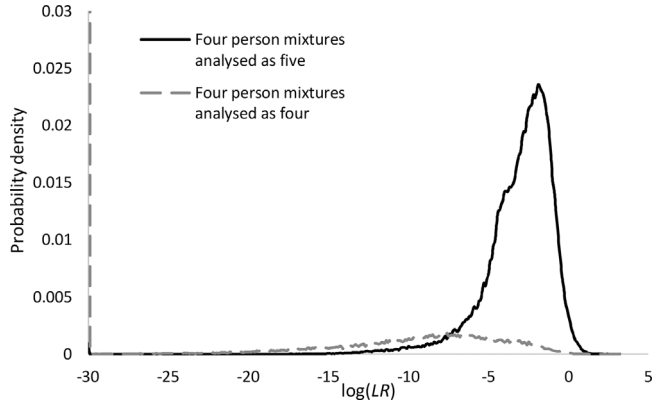

Fig. 9. The *LR*s for H_d true for four person mixtures from one laboratory under the assumption of N and N + 1. 81% of four person mixtures analysed as four resulted in $LR = 0$, represented by $\log_{10}(LR) = -30$.

(N + 1) contributors. Fig. 9 shows the results of the same analysis but when considering four person mixtures as originating from either four (N) or five (N + 1) individuals. The bulk of the distribution for the three person mixtures analysed as three is at $LR = 0$ (90% of all *LR*s) represented by $\log_{10}(LR) = -30$ in Fig. 8. In Fig. 9, 81% of four person mixtures analysed as four resulted in $LR = 0$, again represented by $\log_{10}(LR) = -30$.

Figs. 8 and 9 show that, when analysed using the true number of contributors, the instances of H_d true comparisons that lead to outright exclusions is greatly increased. Put another way, inflating the number of contributors leads to an increase in non-zero *LR*s. In fact, the most common occurrence from inflating the number of contributors is that during deconvolution the additional proposed contributor is assigned a very low template (near 0) and can possess any genotype (including complete dropout) with relatively even weight. This is visually seen in Figs. 8 and 9 by the peak of $\log_{10}(LR)$s just below 0.

3.4. Allele sharing

A demonstration of the effect that allele sharing has on the *LR* is confounded by other factors that affect the magnitude of the *LR*, such as:

- The amount of DNA that the individual has donated to the sample,
- The mixture proportions of the contributors (mixtures at an even mixture proportion will tend to have lower *LR*s, due to the reduction in information that peak heights provide to determine genotype sets),
- Masking of minor contributors in stutter positions of major contributors.

An individual that shares 100% of alleles with the other contributors to a mixture can still have their genotype resolved completely, based on peak heights, given the right circumstances (as seen in Fig. S8 for the family set). The ability to use peak heights in this way is one of

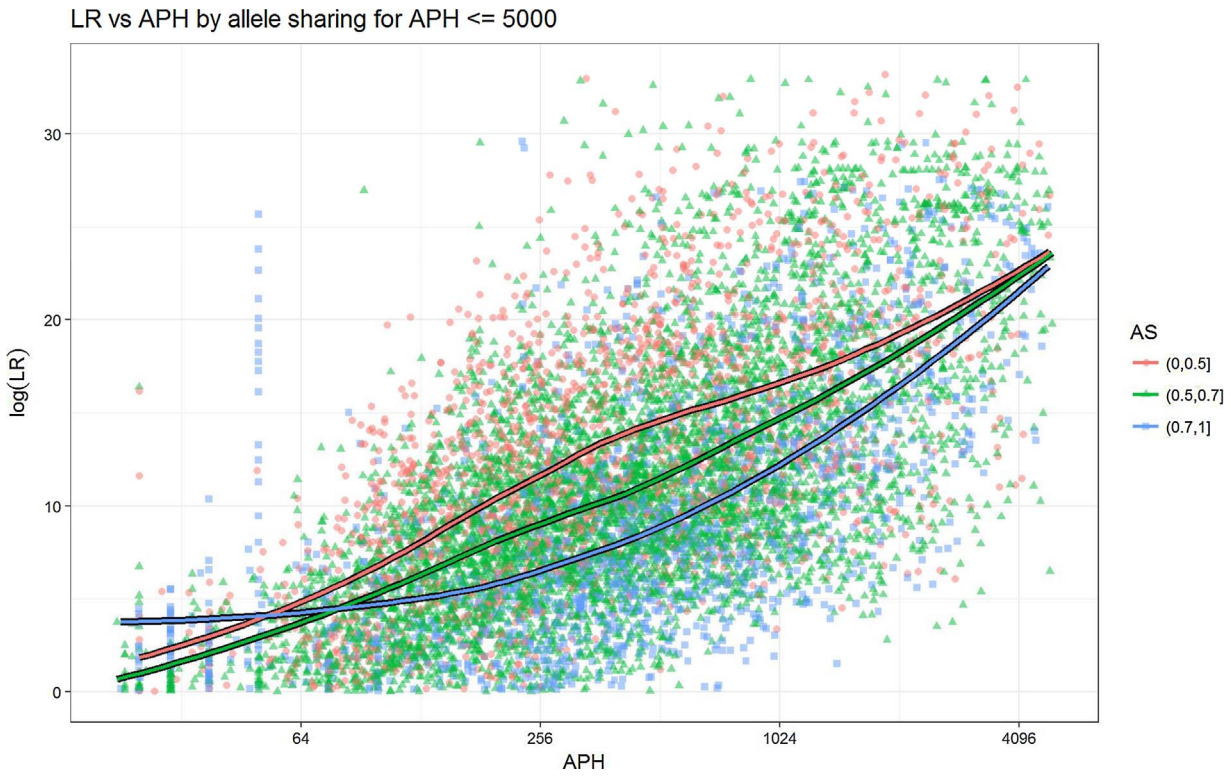

Fig. 10. The size of the $\log_{10}(LR)$ by considering differing amounts of input DNA (*APH*) and amount of allelic sharing (AS). The set of data points with high AS (0.7,1] are a family set (father, mother, children) where all alleles from the children are masked by the parents and therefore *APH* was set to half of the AT.

the main drivers for the differences in *LR*s produced between fully and semi-continuous systems. In Fig. 10 we show the *LR* (on $\log_{10}$ scale) for all data in the study, broken up into three categories of allele sharing, 0 to 0.5, 0.5–0.7 and 0.70–1.0. The lines in Fig. 10 are LOWESS lines to demonstrate the general trends of the data.

From Fig. 10, it appears that the greater the allele sharing, the less the power there is to discriminate a true contributor from a non-contributor. This trend is intuitive as it would be expected that the more an individual's alleles are already accounted for by others in the mixture, the less 'need' there is for someone possessing those alleles to reasonably explain the observed peaks in the mixture. However, further experimentation shows that this apparent trend is totally confounded by the number of contributors to the mixture. Fig. 11 shows the same style of result as Fig. 10, but plotted by number of contributors. In Fig. 11 the recovered weight of evidence is plotted, that is, $\log_{10}(LR)/\log_{10}(1/\text{RMP})$. RMP is the conditional match probability following the Balding and Nichols model [25] and a theta (F_{ST}) of 0.01. Carrying out this transformation accounts for the different profiling systems that are being combined in this *meta*-analysis. In these plots the *y*-axis is bounded by one demonstrating that the *LR* cannot exceed one divided by the random match probability.

The trend seen in Fig. 2 is that higher order mixtures tend to have true contributors that share more alleles (because there are more of them to potentially share), and Figs. S1–S9 demonstrate that higher order mixtures tend to have less discrimination power. Therefore, there is a correlation between allele sharing and *LR* evident in Fig. 10, particularly at low *APH*. In Fig. 11 this trend disappears, showing that it is an effect of number of contributors, and not allele sharing, that is the main driver to *LR* change.

In Fig. 12 we plot a density plot of $\log_{10}(LR)/\log_{10}(1/\text{RMP})$ by the amount of allele sharing of the non-contributors with the true contributors. The $\log_{10}(LR)/\log_{10}(1/\text{RMP})$ cannot exceed one, which would indicate a fully resolved component. Inspection of Fig. 12 shows that as the fraction of shared alleles increases the $\log_{10}(LR)/\log_{10}(1/\text{RMP})$ for the non-contributor increases. As allele sharing of the non-contributors with the true contributors decreases, the $\log_{10}(LR)/\log_{10}(1/\text{RMP})$ decreases with more observations around zero, indicated by the broadening of shape. Fig. 12 shows that non-contributors are unlikely to yield large *LR*s even if they share many alleles with the true contributors. In other words, non-contributors that share most of their alleles with the mixture's donors can typically still be excluded because the peak heights make their inclusion unlikely.

On the other hand, Fig. 6 shows that true contributors can yield *LR*s close to the inverse of the single source match probability even in five person mixtures. This means that at least this mixture donor's component is almost fully resolved on the basis of peak heights. This may be expected, for instance, in a 10:1:1:1:1 mixture where the major may be clearly resolved by simply 'eyeballing' the electropherogram.

4. Discussion

4.1. Performance of the system with regards to contributor number

In principle, we observe less discriminatory *LR*s for true and non-contributors when the number of assigned contributors increases. This has been demonstrated previously using STRmix™ [14,21]. This does not mean that mixed DNA profiles containing more contributors are less reliable, just that they are less informative with respect to potential contributors.

The *true* number of contributors to a crime profile is never known. Within this work we have used the apparent number of contributors when interpreting the mixtures. Apparent N was determined by each submitting laboratory using their own validated methods. The assigned N can be fewer than the true N when individuals within a profile have "dropped out" (their alleles falling below the detection limit of the CE) and within mixtures of contributors with high amounts of allele sharing (an extreme example being mixtures of related individuals). Apparent N may be assigned a number higher than true N in the presence of artefacts, such as stutter, that are larger than expected. This assignment can be confounded in saturated profiles.

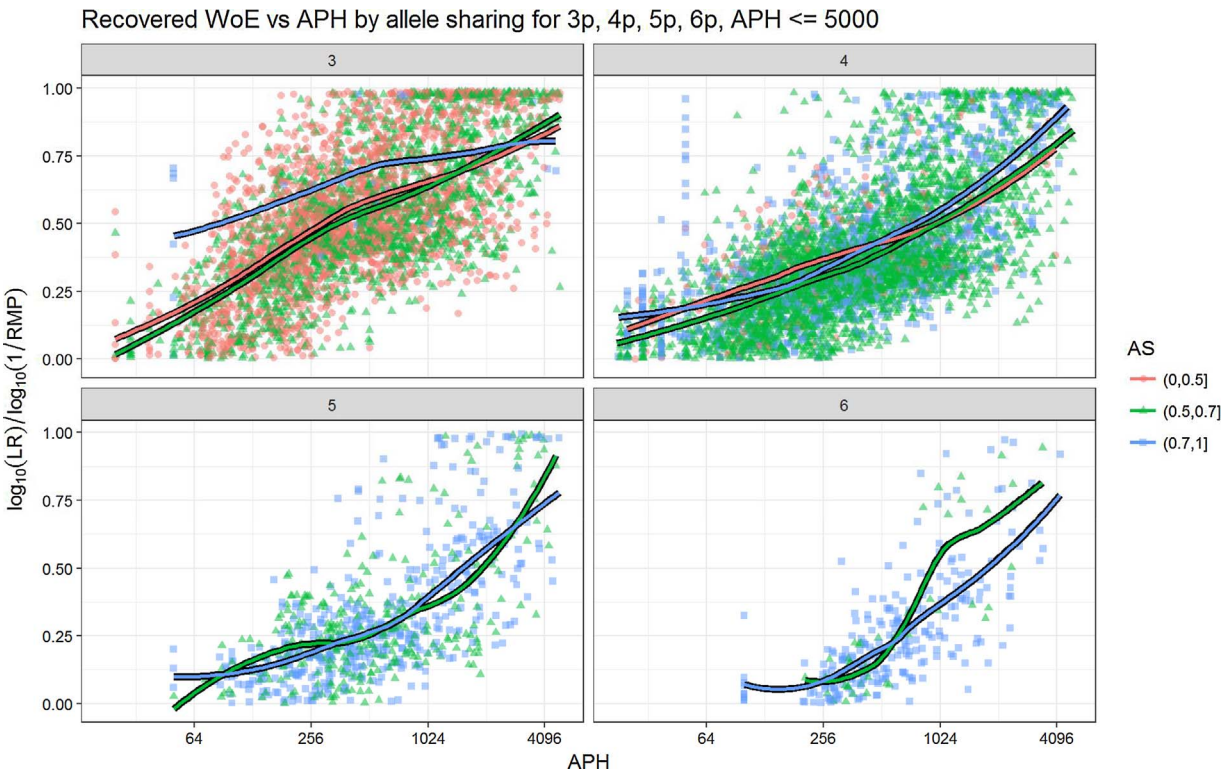

Fig. 11. The size of the recovered weight of evidence $\log_{10}(LR)/\log_{10}(1/\text{RMP})$ by considering differing amounts of input DNA (*APH*) and amount of allelic sharing (AS) plotted by true number of contributors.

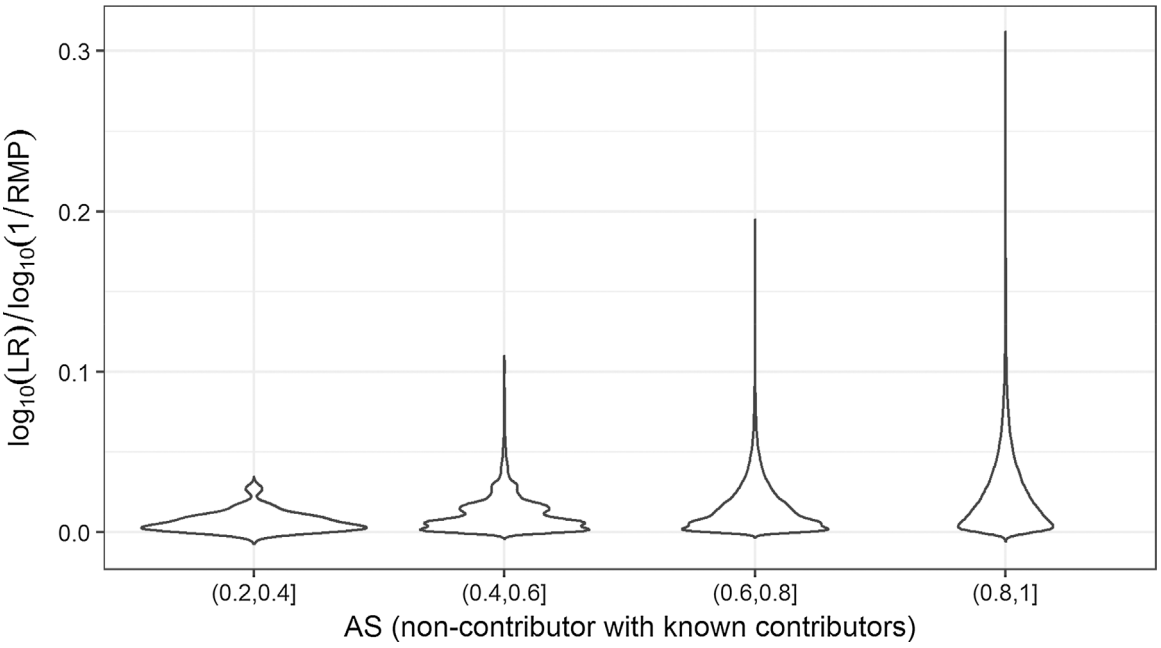

Fig. 12. Density plot of $\log_{10}(LR)/\log_{10}(1/\text{RMP})$ by the amount of allele sharing of the non-contributors with the true contributors.

As the number of contributors to a DNA profile increases, the DNA mixture becomes more complex. Figs. S1 through S9 show *LR*s generated for H_p and H_d true for apparent three, four and five person mixtures plotted against *APH*. As the number of contributors to the mixture increases the *LR*s trend towards one. This holds true for both H_p and H_d true although the effect for H_d true data is less clear given the number of data. As the number of contributors to a mixture increases, so too do the potential genotype combinations that can explain the observed data. This results in an overall reduction in the weights assigned to each genotype set, as these weights are spread across more potential genotype sets. This behaviour was previously described by Taylor [21].

When overestimating the number of contributors to a mixture (N + 1) the *LR* generally decreased for true contributors. This can be explained by STRmix™ spreading the weights for the true donors across more genotype sets. For four person mixtures the magnitude of the effect on the *LR* for known contributors was somewhat dependent on the proportion that the donor contributed to the mixture. The effect was greater for minor contributors to the mixture and less for major contributors (represented by more data points on the $x = y$ line within Fig. 7). Overestimating the number of contributors had little or no effect on the *LR* of the major contributor to the mixture, demonstrated by the largest circles sitting on the $x = y$ trend line. In these cases the additional proposed contributor was modelled as a trace contributor, sharing alleles with the true minor contributors to those mixtures and having little effect on the major. For the three person mixtures the effect was more visible across a range of mixture proportions. This was likely due to similarities in mixture proportions of the different contributors, with no obvious major contributors.

The effect of overestimation of the number of contributors was also determined for non-contributors using H_d true tests. When assuming N + 1 the number of occurrences of non-contributors resulting in non-exclusionary *LR*s increased. During deconvolution the additional

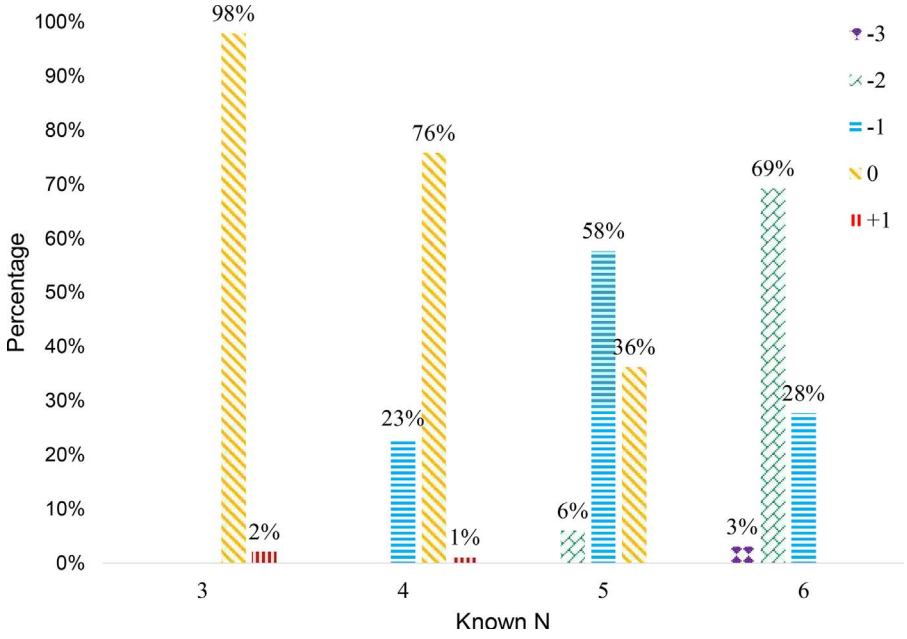

Fig. 13. Plot of percentage of mixtures showing various differences between apparent N and known N against known N. As an example, −1 indicates apparent N was one fewer than known N.

proposed contributor is assigned very low template and can possess any genotype leading to these results.

In summary, overestimation of the number of contributors generally leads to lower *LR*s for true contributors (Fig. 7) and an increase in *LR*s for non-contributors (Fig. 8).

Underestimating the number of contributors can result in false exclusions of true donors. In this study, this is seen when apparent N is fewer than true N. This is demonstrated in the H_p true plots within the Supplementary material where apparent N that differs from known N are indicated with a different plotting symbol.

When assigning N, for false donors the only risk is overestimation, as there is a small increase in the number of very low grade false inclusions. With respect to the *LR* for true donors, you are either correct or conservative when N is either under or overestimated.

In Fig. 13 we provide a plot showing the level of over and underestimation of the apparent N compared to the known N in this study.

Fig. 13 shows that an underestimation of N was more common than an overestimation of N. There are three broad reasons why N might be underestimated:

1) One contributor has donated so little DNA that their presence is unseen in the DNA profile, we call this the tiny minor scenario;
2) Contributors are present so that one or more is completely masked by others in the profile, and in a way so that peak height does not reveal their presence. This is the hidden contributor scenario;
3) There is a combination of multiple low-level contributors that, due to some masking and some dropout, produce a profile where the apparent number of contributors is fewer than the known number of contributors. This is the low level donors' scenario.

Each of these is discussed in turn below.

4.1.1. The tiny minor

Any profile is a result of fragments of DNA that have been aliquoted from a DNA extract and then amplified during PCR. There exists a possibility that no DNA fragments from a minor DNA donor have been sampled for PCR. We first ask what we consider to be the correct number of contributors; the number of different individual's DNA in the DNA extract, or the number of different individual's DNA in the PCR? If it is the former, then we would ask; if the individual has contributed so little DNA that the observed fluorescence in the DNA profile is not affected by their presence, then what purpose is served by considering them as a contributor? We note that many of the underestimates of number of contributors in this study arise from such situations.

4.1.2. The hidden contributor

Consider a DNA profile where multiple individuals, are contributing to a DNA profile, however they possess sufficient allelic overlap so that the DNA profile appears as a lower order mixture. The apparent number of contributors being lower than the known number of contributors relies on the DNA profile being formed in such a way that peak imbalances will not indicate the true number of contributors. For example, a combination of two individuals who are homozygous at each locus, combined in equal proportions to a DNA sample will always appear single source. However, this risk of multiple contributors being combined to meet theses specifications is very remote, and artificial. It only tends to occur in mixtures of family members, such as a child and their parents donating equal amounts of DNA to a sample. The Coble et al. [26] experiment is valuable but does not take into account peak heights, and so the study does not reflect the information that peak heights provide analysts in their assignment of N. This is evident in the difference between the results obtained by Coble et al. and our work. For example, Coble et al. reported the probability of a known five-person mixture presenting as an apparent five person mixture was less than 0.01, whereas in our study, based on human assignment, this probability is 0.36 (and noting that many of the remaining mixtures fall into the tiny minor and low donor scenarios).

4.1.3. The low level donors' scenario

This scenario is where there are multiple low level contributors, who are present in low amounts such that they exhibit significant dropout and so in combination the apparent number of contributor is fewer than the known number of contributors. This is a scenario that could plausibly occur with reasonable probability when multiple low level contributors are present (see [16] for an exploration of this). Experimentation has shown that very low level contributors will yield *LR*s of approximately one. It is likely that when analysed under the known number of contributors, all true (and a majority of false) contributors give this neutral *LR* value. In other words, the profile does not have the information in order to distinguish true from false donors. If analysed as the apparent number of contributors then the likely outcome is an exclusion of the known contributors (and more exclusions of non-contributors). The primary difference in *LR* between known and apparent number of contributors is between neutral and possibly exclusionary, which we could argue presents less risk of misleading a court.

4.1.4. Overestimating the number of contributors

Our studies show that the chance of overestimating N in relation to the known value is less common that underestimation and cannot be predicted so easily by simulation as in Coble et al. [26]. It requires two events to occur:

1) There is a stochastic event, such as a peak imbalance, high stutter or drop-in, which occurs at an improbable level,
2) The analyst interpreting the profile feels that the out-of-place fluorescence has resulted in a profile that is more likely to exist if it has originated from more contributors that the known number of contributors.

Fig. 7 shows that the effect of overestimation of N is relatively mild on known contributors to a DNA profile. STRmix™ assigns near-zero mass to the non-existent contributor, leaving the other contributors relatively unchanged. The largest effect is to decrease the *LR* for minor known contributors. For non-contributors, Fig. 8 shows the effect that has previously been described, i.e. that an overestimation of N tends to increase low-level *LR*s for non-contributors. In effect the experiment is showing the practical functioning of the catch-all statement suggested earlier.

Our findings show that as mixture complexity increases, the ability of an analyst to designate the known number of contributor is reduced. As explained, it is actually often the apparent number of contributors that is the more appropriate value to choose for analysis. In assigning apparent number of contributors the overwhelming result is alignment with the desired trends in *LR*s with regards to profile complexity and DNA amount (i.e. those described in [21], where known number of contributors was used for all analyses) are obtained. In the rare circumstances where the known contributors were not supported as donors of DNA to the profile, this was due to one of the three underestimate conditions described above in 4.1.1 through 4.1.3 above.

4.2. Performance as a function of amount of allele sharing

Within Fig. 10 the trend is that the greater the allele sharing, the less the power to discriminate a true contributor from a non-contributor. However, this relationship is dominated by the number of contributors within the mixture (as seen in Fig. 11). Higher order mixtures result in less informative *LR*s. This effect is related more to the number of contributors within a mixture than the amount of allele sharing between contributors within the mixture. There is a relationship between the number of contributors and proportion of allele sharing within a mixture. It has previously been shown that the probability of a higher order mixture appearing as having originated from one fewer individual

based on allele count alone is high [26,27]. For example, Coble et al. calculated the probability of a six contributor profile appearing as a five contributor profile based on allele count as 0.8599 for the GlobalFiler™ 24 locus multiplex [26]. The study by Coble et al. did not take into account peak height, thereby making the values in their study a worst case scenario.

4.3. Performance of the system with regards to amount of DNA

In principle, we observe less discriminatory *LR*s for true and non-contributors when the *APH* (template) decreases per contributor. Again, this does not mean that mixed DNA profiles with contributors containing less DNA are unreliable, just they are less informative with respect to the true and non-contributors.

PCAST describe limits on PG reliability based on mixture proportion and number of contributors. Per contributor template is more informative of *LR* than mixture proportion. With respect to mixture proportion, the limit is not the software but the hardware. For example, assuming a minor contributor's alleles within a mixture are present just above the analytical threshold of a 3130 (typically 50 rfu) and a major contributor's alleles are at the saturation limit (typically 7000 rfu), this would be maximum mixture proportion of 140:1. 2293 out of the 2825 submitted profiles had at least one component who contributed less than 20% of the sample.

5. Conclusion

In their review of published literature validating probabilistic genotyping, PCAST surmised that the limits of foundational validity extended to three person mixtures where the person of interest made up at least 20% of the profile. What was not taken into account during the PCAST review was a wealth of unpublished validation material residing in laboratories that had validated (or were in the process of validating) probabilistic genotyping software. Due to our involvement with STRmix™ we are aware of the breadth of such validation material for STRmix™ specifically, and assume that similar material must be present for other probabilistic genotyping systems. A disconnect exists between the PCAST desire for laboratories to publish their validation material in peer reviewed journals and the general resistance to such publications by the journals themselves. This is for the completely understandable reason that they are generally not novel, or, individually, of general interest to the forensic community.

> *PCAST has said "When further studies are published, it will likely be possible to extend the range in which scientific validity has been established to include more challenging samples. As noted above, such studies should be performed by or should include independent research groups not connected with the developers of the methods and with no stake in the outcome*."

There has already been an example of published material that extend the PCAST limits, from the Forensic Biology laboratory at the Federal Bureau of Investigation [14]. We add to that published work, by compiling the STRmix™ validation material from 31 laboratories, which allows a novel look at data spanning laboratory technology and process. PCAST highlighted four key areas that they felt additional validation would be merited:

(1) How well does the method perform as a function of the number of contributors to the mixture? How well does it perform when the number of contributors to the mixture is *unknown*?
(2) How does the method perform as a function of the number of alleles shared among individuals in the mixture? Relatedly, how does it perform when the mixtures include related individuals?
(3) How well does the method perform—and how does accuracy degrade—as a function of the absolute and relative amounts of DNA from the various contributors?
(4) Under what circumstances—and why—does the method produce results (random inclusion probabilities) that differ substantially from those produced by other methods?

We address points 1 to 3 in this study. It is unknown whether further addendums will be released by the PCAST group, or whether there are any plans for a follow-up study in the future. The material we provide here demonstrates a foundational validity of, at least, the STRmix™ software method for complex, mixed DNA profiles to levels well beyond the complexity and contribution levels suggested by PCAST. The study was done in accordance with the specific manner outlined in the PCAST report.

Acknowledgements

This work was supported in part by grant 2011-DN-BX-K541 from the US National Institute of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of their organisations. The authors would like to thank Professor James Curran for his help in creating the plots in Fig. 1.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2018.01.003.

References

[1] President's Council of Advisors on Science and Technology, PCAST Releases Report on Forensic Science in Criminal Courts, (2016).
[2] President's council of advisors on science and technology, An Addendum to the PCAST Report on Forensic Science in Criminal Courts, (2016).
[3] A.A. Mitchell, J. Tamariz, K. O'Connell, N. Ducasse, Z. Budimlija, M. Prinz, et al., Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in, Forensic Sci. Int. Genet. 6 (2012) 749–761.
[4] M.W. Perlin, M.M. Legler, C.E. Spencer, J.L. Smith, W.P. Allan, J.L. Belrose, Validating TrueAllele® DNA mixture interpretation, J. Forensic Sci. (2011) 2011.
[5] J.-A. Bright, D. Taylor, C.E. McGovern, S. Cooper, L. Russell, D. Abarno, et al., Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles, Forensic Sci. Int. Genet. 23 (2016) 226–239.
[6] F.M. Götz, H. Schönborn, V. Borsdorf, A.-M. Pflugbeil, D. Labudde, GenoProof Mixture 3—New software and process to resolve complex DNA mixtures, Forensic Sci. Int. Genet. Suppl. Ser. (2017).
[7] S. Manabe, C. Morimoto, Y. Hamano, S. Fujimoto, K. Tamaki, Development and validation of open-source software for DNA mixture interpretation based on a quantitative continuous model, PLoS One 12 (2017) e0188183.
[8] Scientific Working Group on DNA Analysis Methods (SWGDAM), Guidelines for the Validation of Probabilistic Genotyping Systems, (2015).
[9] M.D. Coble, J. Buckleton, J.M. Butler, T. Egeland, R. Fimmers, P. Gill, et al., DNA Commission of the International Society for Forensic Genetics: recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications, Forensic Sci. Int. Genet. 25 (2016) 191–197.
[10] D. Taylor, J.-A. Bright, J. Buckleton, The interpretation of single source and mixed DNA profiles, Forensic Sci. Int. Genet. 7 (2013) 516–528.
[11] J.-A. Bright, D. Taylor, J.M. Curran, J.S. Buckleton, Developing allelic and stutter peak height models for a continuous method of DNA interpretation, Forensic Sci. Int. Genet. 7 (2013) 296–304.
[12] The New York City Office of Chief Medical Examiner, Internal Validation of STRmix™ V2.4 for Fusion NYC OCME, (2016).
[13] District of Columbia Department of Forensic Science Forensic Science Laboratory Forensic Biology Unit. Internal Validation of STRmix™ V2.3, (2015).
[14] T.R. Moretti, R.S. Just, S.C. Kehl, L.E. Willis, J.S. Buckleton, J.-A. Bright, et al., Internal validation of STRmix for the interpretation of single source and mixed DNA profiles, Forensic Sci. Int. Genet. 29 (2017) 126–144.
[15] T.R. Moretti, L.I. Moreno, J.B. Smerick, M.L. Pignone, R. Hizon, J.S. Buckleton, et al., Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States, Forensic Sci. Int. Genet. 25 (2017) 175–181.
[16] D. Taylor, J. Buckleton, J.-A. Bright, Does the use of probabilistic genotyping change the way we should view sub-threshold data, Aust. J. Forensic Sci. 49 (2017) 78–92.
[17] D. Taylor, J. Buckleton, I. Evett, Testing likelihood ratios produced from complex DNA profiles, Forensic Sci. Int. Genet. 16 (2015) 165–171.
[18] J.-A. Bright, D. Taylor, J. Curran, J. Buckleton, Searching mixed DNA profiles directly against profile databases, Forensic Sci. Int. Genet. 9 (2014) 102–110.
[19] J.-A. Bright, J.M. Curran, J.S. Buckleton, The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation, Forensic Sci. Int.

Genet. 12 (2014) 208–214.
[20] J.L. Hintze, R.D. Nelson, Violin plots a box plot-density trace synergism, Am. Stat. 52 (1998) 181–184.
[21] D. Taylor, Using continuous DNA interpretation methods to revisit likelihood ratio behaviour, Forensic Sci. Int. Genet. 11 (2014) 144–153.
[22] H. Wickham, Ggplot2–Elegant Graphics for Data Analysis, 2nd edition, Springer-Verlag, New York, 2016.
[23] Applied Biosystems User Bulletin Applied Biosystems® 3500/3500xL Genetic Analyzer, Life Technologies internal report, Foster City, CA), 2011.
[24] D. Taylor, J. Buckleton, J.-A. Bright, Factors affecting peak height variability for short tandem repeat data, Forensic Sci. Int. Genet. 21 (2016) 126–133.
[25] D.J. Balding, R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, Forensic Sci. Int. 64 (1994) 125–140.
[26] M.D. Coble, J.-A. Bright, J.S. Buckleton, J.M. Curran, Uncertainty in the number of contributors in the proposed new CODIS set, Forensic Sci. Int. Genet. 19 (2015) 207–211.
[27] J.S. Buckleton, J.M. Curran, P. Gill, Towards understanding the effect of uncertainty in the number of contributors to DNA stains, Forensic Sci. Int. Genet. 1 (2007) 20–28.