Forensic Science International: Genetics 40 (2019) 150–159

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Interpreting a major component from a mixed DNA profile with an unknown number of minor contributors

Todd Bille[a,*], Steven Weitz[a], John S. Buckleton[b,c], Jo-Anne Bright[b]

[a] Bureau of Alcohol, Tobacco, Firearms, and Explosives Laboratory, National Laboratory Center, 6000 Ammendale Road, Beltsville, MD, 20705, United States
[b] Institute of Environmental Science and Research Limited, Private Bag 92021, Auckland, 1142, New Zealand
[c] University of Auckland, Department of Statistics, Private Bag 92019, Auckland, New Zealand

ARTICLE INFO

Keywords:
Forensic DNA
Probabilistic genotyping
Mixture interpretation
STRmix™

ABSTRACT

Modern interpretation strategies typically require an assignment of the number of contributors (*N*) to a DNA profile. This can prove to be a difficult task, particularly when dealing with higher order mixtures or mixtures where one or more contributors have donated low amounts of DNA. Differences in the assigned *N* at interpretation can lead to differences in the likelihood ration (*LR*). If the number of contributors cannot reasonably be assigned, then an interpretation of the profile may not be able to be progressed.

In this study, we investigate mixed DNA profiles of varying complexity and interpret them altering the assigned *N*. We assign *LR*s for true- and non- contributors and compare the results given different assignments of *N* over a range of mixture proportions. When a component of a mixture had a proportion of at least 10%, a ratio of at least 1.5:1 to the next highest component, and a DNA amount (as determined by STRmix™) of at least 50 rfu, the *LR* of the component for a true contributor was not significantly affected by varying *N* and was therefore suitable for interpretation and the assignment of an *LR*. *LR*s produced for minor contributors were found to vary significantly as the assigned *N* was changed. These heuristics may be used to identify profiles suitable for interpretation.

1. Introduction

In 2006, the International Society for Forensic Genetics (ISFG) published recommendations for the interpretation of forensic DNA mixtures with low-level contributors [1]. They recommended the likelihood ratio (*LR*) as the preferred method of interpretation. The *LR* is a ratio of two probabilities that evaluates the evidence given two or more mutually exclusive propositions:

$$LR = \frac{\Pr(E|H_1)}{\Pr(E|H_2)}$$

Where *E* is the DNA profile. In forensic DNA typing, H_1 will typically consider that the person of interest (POI) is a contributor to the recovered DNA, whilst H_2 considers that the DNA originates from an unknown individual(s).

More recently, a number of papers have been published describing methods for the probabilistic genotyping (PG) of mixtures and assignment of *LR*s [2–5]. Probabilistic genotyping refers to the use of biological models and statistical theory to calculate *LR*s and/or infer genotypes for the DNA typing results of forensic samples [6].

Nearly all PG software programs currently used in casework require the manual assignment of the *N* by the analyst prior to interpretation of the mixed DNA profile. *N* is described as a nuisance parameter in profile interpretation; a value that is not of immediate interest but that must be considered in the model [7]. The *N* may be the same under both H_1 and H_2 (constrained) or allowed to differ under each proposition (unconstrained) [8]. Assigning *N* can be challenging, particularly with low-level components [9,10], and in higher order mixtures [11]. A number of different methods have been described to help with this assessment [12–16]. Taylor et al. describe a probabilistic approach that chooses values for *N* that maximise the probability of the observed profile under both the numerator and denominator of the *LR* [17].

In our experience, most laboratories assign *N* using the Maximum Allele Count (MAC) method [14,15] whilst also considering peak height information. A common practice is that if the number of contributors cannot be assigned with a reasonable degree of confidence, then an interpretation of the profile will not be undertaken. This may be the case where there are multiple low-level contributors with indications of allelic dropout or where dealing with higher order mixtures, for example those that appear to originate from four or more contributors.

* Corresponding author.
E-mail address: todd.bille@atf.gov (T. Bille).

https://doi.org/10.1016/j.fsigen.2019.02.017
Received 27 December 2018; Received in revised form 4 February 2019; Accepted 20 February 2019
Available online 22 February 2019
1872-4973/ Published by Elsevier B.V.

Table 1
Summary of constructed mixed DNA profiles including mixture proportions, number of PCR replicates per profile, target template amount, total number of interpretations per sample, and assigned *N* per interpretation. A total of 147 interpretations were undertaken.

Target *N*	Mixture proportions	Template DNA (ng)	PCR replicates	*N* tested	Number of interpretations
5	32:32:32:2:2	0.5, 1.0	2	4,5	8
4	32:32:32:4	0.5, 1.0	2	3,4	8
4	33:32:32:3	0.5, 1.0	2	3,4	8
4	33:33:33:1	0.5, 1.0	2	3,4	8
5	48:47:2:2:1	0.5, 1.0	2	3,4,5	12
4	48:48:2:2	0.5, 1.0	2	3,4	8
3	48:48:4	0.5, 1.0	2	3,4	8
5	75:20:2:2:1	0.25, 0.5, 1.0	2	3,4,5	18
4	75:20:3:2	0.25, 0.5, 1.0	2	3,4	12
3	75:20:5	0.25	2	2,3	4
4	85:5:5:5	0.5, 1.0	2	3,4,5	11*
4	85:5:5:5	0.25	2	2,3,4	6
5	96:1:1:1:1	0.5, 1.0	2	3,4,5	12
4	97:1:1:1	0.5, 1.0	2	3,4,5	12
3	98:1:1	0.5, 1.0	2	2,3,4	12

* note that one of the 85:5:5:5 1.0 ng replicates could not be explained assuming three contributors, therefore only 11 interpretations were progressed.

Some work has been undertaken investigating the effect on the *LR* of a misassignment in *N* [9,10,16,18,19]. Generally, underassigning *N* will result in false exclusions of known (true) contributors. If fewer contributors to a mixture are assumed compared to the ground truth, less allele sharing is permitted and genotypes may be incorrectly assigned leading to the false exclusion of true contributors. Overassigning *N* will typically increase the number of false inclusions with low *LR* values (referred to as low-grade adventitious matches), and may lower the *LR* for true contributors. This is because more genotypes must be considered, and the probability of the evidence given the genotypes (termed weights) will be spread across the range of genotypes. The *LRs* for major contributors to DNA profiles are less likely to be affected by an overassignment of *N*.

Table 2
Summary of H_2 true *LRs* greater than 1. The largest *LR* for a non-contributor was 496.

LR Range	# of Non-Contributors	% of Total H_2 Comparisons
$1 \leq LR < 2$	2750	1.867%
$2 \leq LR < 10$	1605	1.090%
$10 \leq LR < 50$	196	0.133%
$50 \leq LR < 100$	15	0.010%
$100 \leq LR < 200$	5	0.003%
$200 \leq LR < \infty$	5	0.003%
Total	4576	3.107%

In some situations, it may be possible to calculate an *LR* even if *N* cannot reasonably be assigned due to ambiguity in the number of minor contributors. In profiles where there is a large difference in the contributions from the components of a mixture (mixture proportions), higher level components of the mixture will be unaffected by an increase or decrease in the *N*, but the minor components will be significantly affected. This is generally due to the additional contributor being considered as a minor contribution (with low mixture proportion)

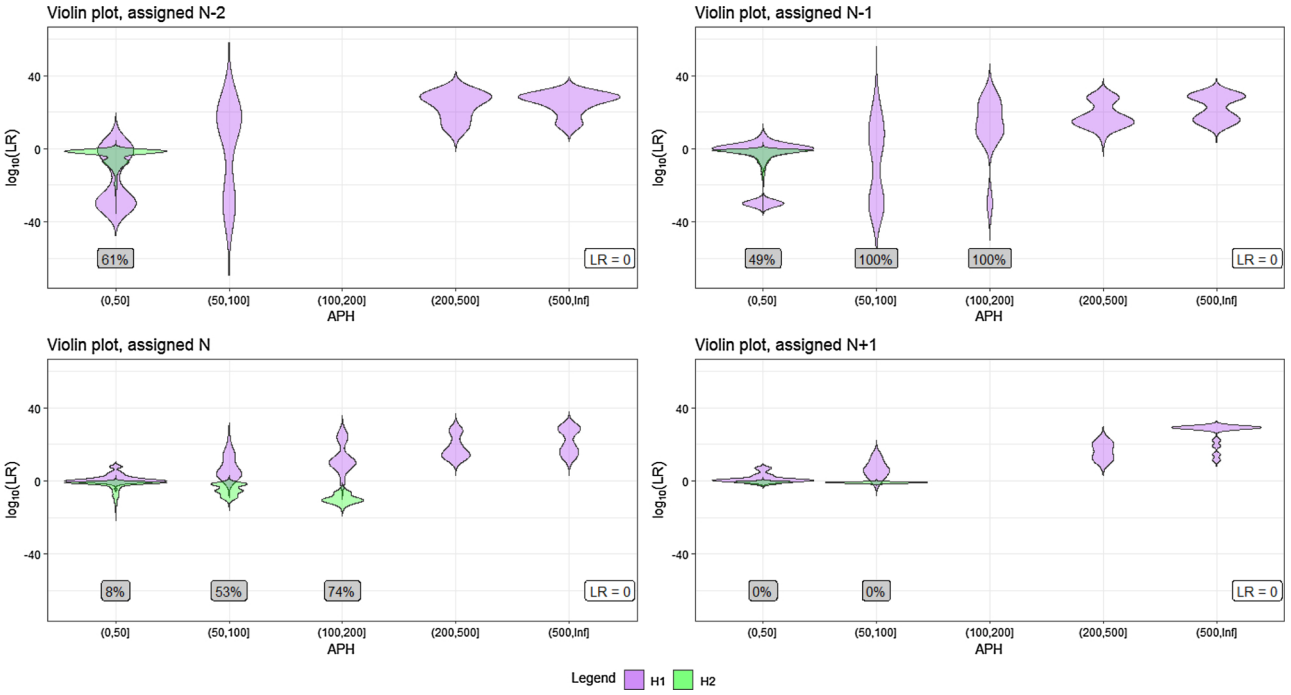

Fig. 1. Violin plots of $\log_{10}(LR)$ versus APH (in rfu) for all mixtures separated by assigned *N* relative to target *N* for all 147 interpretations (614 H_1 true *LRs* and 147,268 H_2 true). *LRs* for known contributors are plotted in purple (H_1 true) and *LRs* for non-contributors plotted in green (H_2 true). Exclusions ($LR = 0$) for non-contributors are not plotted and the percentage of data is given at the bottom of each plot (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Table 3
Summary of the amount of allele sharing between the non-contributors with the minor contributor for the twenty highest H_2 true *LR*s (involving 16 mixtures).

Sample	Assigned *N*	Random individual	*LR*	% Allele Sharing with Minor Components Separately	% Allele Sharing with Minor Components Combined
85:5:5:5 0.25 ng rep A	*N*-1	893	496	40 - 48%	71%
98:1:1 0.5 ng rep B	*N*-1	243	465	45 - 48%	62%
98:1:1 0.5 ng rep B	*N*-1	172	346	40 - 48%	60%
85:5:5:5 0.25 ng rep A	*N*	893	296	40 - 48%	71%
48:47:2:2:1 0.5 ng rep A	*N*-2	164	281	33 - 45%	81%
33:33:33:1 1.0 ng rep B	*N*	521	160	55%	55%
98:1:1 0.5 ng rep B	*N*	243	139	45 - 48%	62%
98:1:1 1.0 ng rep A	*N*	629	106	48 - 50%	67%
96:1:1:1:1 0.5 ng rep B	*N*-2	337	102	31 - 45%	74%
98:1:1 0.5 ng rep B	*N*	172	101	40 - 48%	60%
98:1:1 0.5 ng rep B	*N*-1	388	99	38 - 48%	62%
48:47:2:2:1 0.5 ng rep A	*N*-1	164	90	33 - 45%	81%
48:48:4 0.5 ng rep B	*N*+1	926	89	38%	38%
75:20:2:2:1 0.25 rep B	*N*-2	598	83	26 - 50%	79%
96:1:1:1:1 0.5 ng rep A	*N*-2	343	82	33 - 40%	79%
96:1:1:1:1 0.5 ng rep A	*N*-1	343	81	33 - 40%	79%
75:20:2:2:1 0.25 rep B	*N*-1	288	81	24 - 36%	67%
48:48:4 0.5 ng rep B	*N*+1	279	81	48%	48%
96:1:1:1:1 0.5 ng rep A	*N*	343	79	33 - 40%	79%
75:20:2:2:1 0.25 rep B	*N*	288	69	24 - 36%	67%

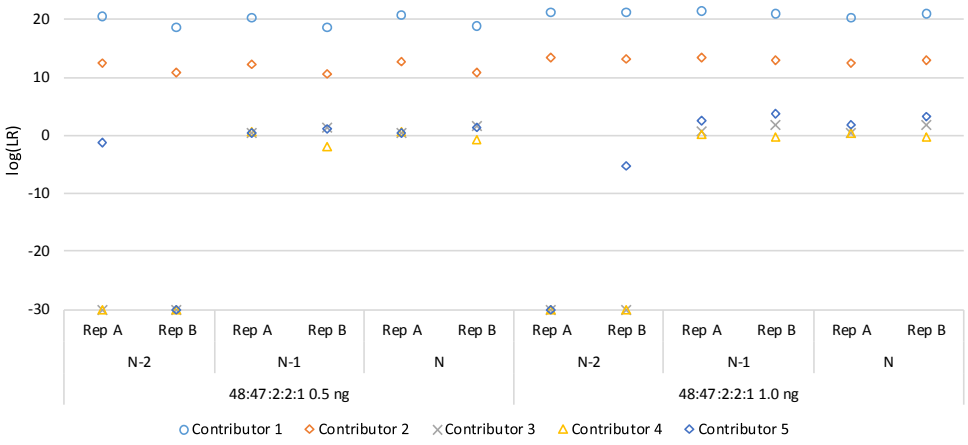

Fig. 2. $\text{Log}_{10}(LR)$ for the five known contributors to two replicate amplifications of a sample with mixture proportions 48:47:2:2:1 ($N = 5$) amplified with 0.5 ng and 1.0 ng total template and interpreted with varying *N*.

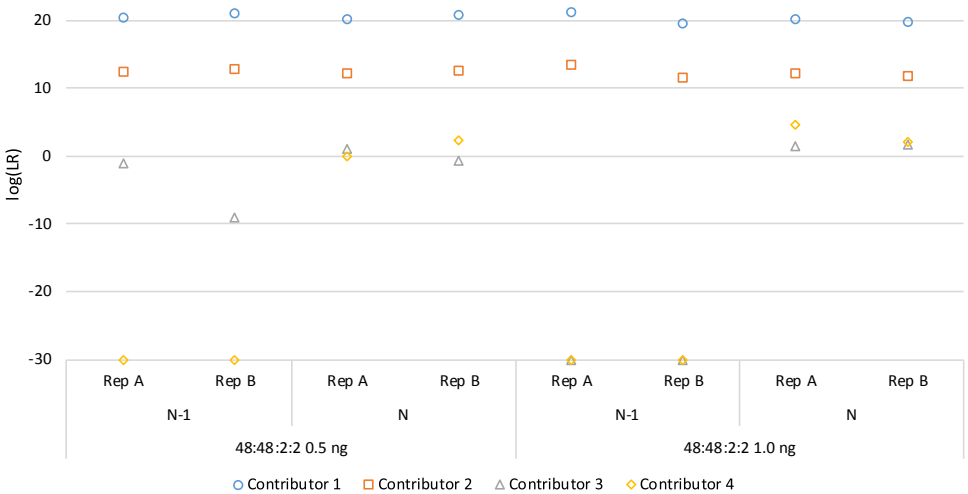

Fig. 3. $\text{Log}_{10}(LR)$ for the four known contributors to two replicate amplifications of a sample with mixture proportions 48:48:2:2 ($N = 4$) amplified with 0.5 ng and 1.0 ng total template and interpreted with varying *N*.

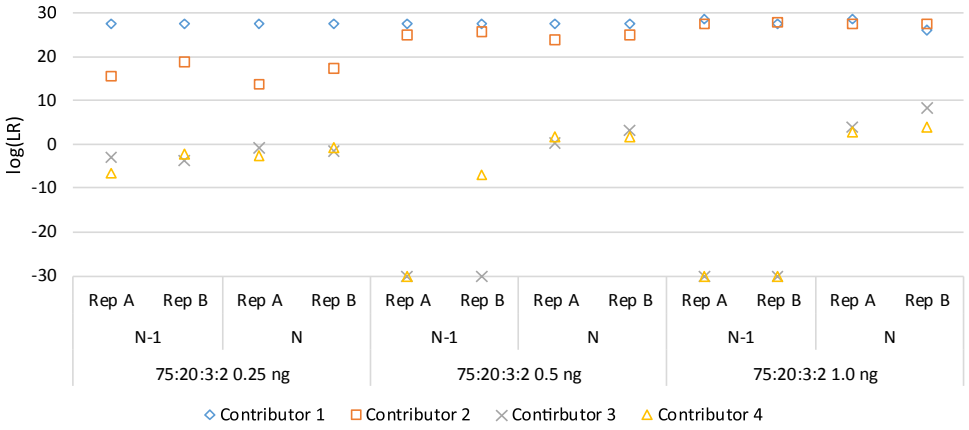

Fig. 4. $Log_{10}(LR)$ for the four known contributors to two replicate amplifications of a sample with mixture proportions 75:20:3:2 ($N = 4$) amplified with 0.25 ng, 0.5 ng, and 1.0 ng total template and interpreted with varying N.

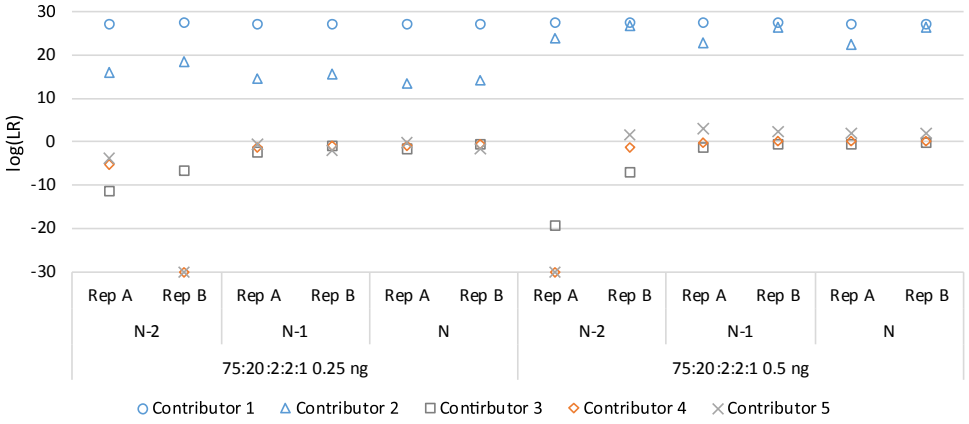

Fig. 5. $Log_{10}(LR)$ for the five known contributors to two replicate amplifications of a sample with mixture proportions 75:20:2:2:1 ($N = 5$) amplified with 0.25 ng and 0.5 ng total template and interpreted with varying N.

Table 4
STRmix™ assigned DNA amounts (rfu) for contributors to replicate 75:20:3:2 mixtures amplified given different template (ng) and assigned N. The ratio of the each major contributors' DNA amount to the sum of the remaining lesser components is also given.

Template (ng)	Assigned N	Replicate	DNA amount 1	DNA amount 2	DNA amount 3	DNA amount 4	Major (0.75): Sum of Minors (contributors 2-4)	Major (0.2): Sum of Minors (contributors 3-4)
0.5	3	A	690	133	42	–	3.9:1	3.2:1
0.5	4	A	719	129	50	21	3.6:1	1.8:1
0.5	3	B	644	179	34	–	3:1	5.3:1
0.25	3	A	256	49	32	–	3.2:1	1.5:1
0.25	4	A	257	46	29	14	2.9:1	1.1:1
0.25	3	B	387	60	39	–	3.9:1	1.5:1
0.25	4	B	396	55	35	13	3.8:1	1.2:1

thereby influencing how the low level components of the mixture are considered. We ask, under what circumstances is it possible to identify a major component that is suitable for interpretation that is unaffected by a change in N?

In the present study, we interpret 60 three-, four-, and five-person mixed DNA profiles prepared using DNA from known individuals, assigning both the target ('correct') and incorrect N, and investigate whether a contributor with a substantially greater mixture proportion can be interpreted despite ambiguity in N. The profiles examined were designed to have extreme differences between the mixture proportions of the donors.

In order to improve consistency between analysts both within and between laboratories, an empirically-based method to determine when a component is suitable for analysis in these situations is desirable. This paper investigates profiles originating from contributors with large differences in mixture proportions with a view to determining conditions where the LR is not significantly affected by changes in N.

Table 5
STRmix™ assigned DNA amounts (rfu) for contributors to replicate 75:20:2:2:1 mixtures amplified given different template (ng) and assigned *N*. The ratio of the each major contributors' DNA amount to the sum of the remaining lesser components is also given.

Template (ng)	Assigned *N*	Replicate	DNA amount 1	DNA amount 2	DNA amount 3	DNA amount 4	DNA amount 5	Major (0.75): Sum of Minors (contributors 2-5)	Major (0.2): Sum of Minors (contributors 3-5)
0.5	3	A	779	172	53	–	–	3.5:1	3.3:1
0.5	4	A	779	173	56	24	–	3.1:1	2.2:1
0.5	5	A	782	185	57	24	7	2.9:1	2.1:1
0.5	3	B	637	133	22	–	–	4.1:1	6.1:1
0.5	4	B	647	137	23	7	–	3.9:1	4.6:1
0.5	5	B	664	133	22	11	2	4.0:1	3.8:1
0.25	3	A	249	54	33	–	–	2.9:1	1.6:1
0.25	4	A	253	49	33	18	–	2.5:1	1.0:1
0.25	5	A	260	44	30	20	10	2.5:1	0.7:1
0.25	3	B	273	49	32	–	–	3.4:1	1.5:1
0.25	4	B	274	47	31	17	–	2.9:1	1.0:1
0.25	5	B	271	44	30	19	8	2.7:1	0.8:1

2. Method

Sixty mixed DNA profiles were prepared by adding extracted DNA from different individuals in varying proportions as outlined in Table 1. All samples were amplified using the GlobalFiler™ STR multiplex (Thermo Fisher Scientific) and separated on a 3130 Genetic Analyzer (Thermo Fisher Scientific). Profiles were analysed in GeneMapper™ *ID-X* V1.4 using empirically determined per dye analytical thresholds (blue 23 rfu, green 38 rfu, yellow 23 rfu, red 27 rfu, and purple 30 rfu).

Profiles were interpreted using STRmix™ V2.5.11 [5,20]. An *LR* was calculated for the true contributors and 1000 non-contributors given the following propositions:

H1. The DNA originated from the person of interest and *N*-1 unknown contributors

H2. The DNA originated from *N* unknown contributors where *N* is the assigned number of contributors (as per Table 1). The 1000 non-contributors were generated artificially by sampling from the NIST Caucasian allele frequencies [21]. All *LR*s were calculated using the NIST Caucasian allele frequencies and $F_{ST} = 0.01$. Profiles were interpreted assuming different numbers of contributors (*N*) to give a total of 147 interpretations (Table 1). Profiles were interpreted assuming the target *N* and, where possible, *N*-2, *N*-1, and *N*+1. In total, there were 614 H_1 true *LR*s and 147,268 H_2 true *LR*s calculated.

3. Results

In Fig. 1, $\log_{10}(LR)$ grouped by average peak height (APH) for all profiles was plotted using the package ggplot2 in R [2,3]. Average peak height (APH) was calculated for each known contributor using unshared alleles only and those not in stutter positions of other contributors. Profiles are divided into ranges of APH by rfu. The APH for non-contributors was taken as the minimum per contributor value for each sample. Therefore, there may not be data for the non-contributors at the higher APH ranges. For example, in the *N*-2 plot, the minimum APH for a known contributor was < 50 rfu hence all of the non-contributor data is plotted in the first APH group.

Results were plotted by assigned *N* relative to target *N*. For example, *N*-2 is two fewer contributors than the target and *N*+1 one more than target *N*. The width of the shaded area of the violin plots represents the proportion of the data located there. *LR*s for known contributors are plotted in purple (H_1 true) and *LR*s for non-contributors plotted in green (H_2 true). Exclusions ($LR = 0$) for non-contributors are not plotted and the percentage of data falling into this class is given at the bottom of each plot. For example, in the *N*-2 plot, 61% of the non-contributors were excluded ($LR = 0$).

On inspection of Fig. 1 some general conclusions can be drawn. The addition of one contributor above target *N* (*N*+1) resulted in no exclusions ($LR = 0$) for non-contributors (H_2 true). When *N*-1 and *N*-2 were assumed, false exclusions of known contributors were observed. *LR*s for true contributors are shown to increase as an individual contributor's APH increases. Conversely, generally low *LR*s were obtained for non-contributors trending towards $LR = 1$ as the APH decreases. This is the expected result [9,10,22]. The multimodal nature of the density plots is most likely the product of combining the results from all the mixtures that included various mixture ratios. The *LR* results vary as the amount of genotype combination ambiguity increases (e.g. 98:1:1, 75:20:1, 32:32:32:4).

One third of the H_2 true comparisons resulted in $LR = 0$ (48,696/147,268). Approximately 64% of H_2 true comparisons resulted in $LR < 1$ (but not 0) and 3% greater than or equal to 1, $LR \geq 1$. A summary of H_2 *LR*s greater than 1 is given in Table 2. The largest *LR* for a non-contributor was 496 (sample 85:5:5:5 0.25 ng, assigned $N = 3$).

The largest 20 *LR*s from non-contributors were reviewed in more detail (Table 3). Each of the non-contributors aligned with a minor contributor position within a mixture, with the highest APH being 45 rfu. Four of the 20 were interpreted with assigned *N*-2, seven of the 20 with *N*-1, seven with *N*, and two with assigned *N*+1. Of the 20 highest *LR*s assigned to non-contributors these originated from 16 interpretations (9 different mixtures with varying *N*) and involved 13 individuals from the database. Table 3 lists the amount of allele sharing between the non-contributor profiles and each of the minor contributors. In almost all cases, the non-contributor shared at least 40% of its alleles with one or more of the minor contributors.

In the following figures, we describe the results of individual samples to highlight the specific effects that varying the assignment of *N* has on the *LR*s for known contributors. Plots of $\log_{10}(LR)$ for all interpretations are given in the supplementary material. For all mixtures analysed, the *LR*s of the major components did not change from supporting inclusion under target *N* to supporting exclusion under a different assigned *N*.

In Fig. 2, we plot $\log_{10}(LR)$ for a five-person mixture with mixture proportions of 48:47:2:2:1 amplified at 0.5 ng and 1.0 ng total. The two replicate profiles were interpreted individually assuming three (*N*-2),

Table 6
STRmix™ assigned values for mixture proportion, DNA amount, and the $\log_{10}(LR)$ along with the calculated ratio to the next highest component for each contributor to replicates of the 75:20:3:2 mixture. Highlighted rows indicate components where the *LR*s were significantly affected by variation in *N*. Exclusions ($LR = 0$) are represented as $\log_{10}(LR) = -30$.

Mixture Replicate *N*	Known Contributor	Mixture Proportion	DNA Amount (rfu)	Ratio To Next Highest Contributor	$\log_{10}(LR)$
75:20:3:2 1.0 ng A *N*-1	1	78	1152	4.27:1	27.68
	2	18	270	4.66:1	28.62
	3	4	58	-	-30
	4	-	-	-	-30
75:20:3:2 1.0 ng A *N*	1	78	1161	4.64:1	27.49
	2	17	250	4.55:1	28.53
	3	4	55	1.83:1	4.07
	4	2	30	-	2.77
75:20:3:2 1.0 ng B *N*-1	1	81	1558	5.19:1	27.75
	2	16	300	4.41:1	27.68
	3	4	68	-	-30
	4	-	-	-	-30
75:20:3:2 1.0 ng B *N*	1	76	1544	4.71:1	27.50
	2	16	328	3.15:1	26.10
	3	5	104	1.93:1	8.30
	4	3	54	-	4.06
75:20:3:2 0.5 ng A *N*-1	1	80	690	5.19:1	27.62
	2	15	133	3.17:1	24.87
	3	5	42	-	-30
	4	-	-	-	-30
75:20:3:2 0.5 ng A *N*	1	78	719	5.57:1	27.42
	2	14	129	2.58:1	23.94
	3	5	50	2.38:1	1.91
	4	2	21	-	0.44
75:20:3:2 0.5 ng B *N*-1	1	75	644	3.60:1	27.61
	2	21	179	5.26:1	25.74
	3	4	34	-	-6.85
	4	-	-	-	-30
75:20:3:2 0.5 ng B *N*	1	76	1539	4.51:1	27.49
	2	17	341	3.34:1	26.16
	3	5	102	1.89:1	7.94
	4	3	54	-	4.05
75:20:3:2 0.25 ng A *N*-1	1	76	256	5.22:1	27.51
	2	15	49	1.53:1	15.44
	3	9	32	-	-6.44
	4	-	-	-	-3.07
75:20:3:2 0.25 ng A *N*	1	74	257	5.59:1	27.35
	2	13	46	1.59:1	13.86
	3	8	29	2.07:1	-2.43
	4	4	14	-	-0.83
75:20:3:2 0.25 ng B *N*-1	1	80	387	6.45:1	27.52
	2	12	60	1.54:1	18.83
	3	8	39	-	-2.30
	4	-	-	-	-3.78
75:20:3:2 0.25 ng B *N*	1	79	396	7.20:1	27.36
	2	11	55	1.57:1	17.24
	3	7	35	2.69:1	-0.86
	4	3	13	-	-1.32

Table 7
STRmix™ assigned values for mixture proportion, DNA amount, and the $\log_{10}(LR)$ along with the calculated ratio to the next highest component for each contributor to replicates of the 75:20:2:2:1 mixture (1.0 ng template DNA). Highlighted rows indicate components where the *LR*s were significantly affected by *N*. Exclusions ($LR = 0$) are represented as $\log_{10}(LR) = -30$.

Mixture Replicate *N*	Known Contributor	Mixture Proportion	DNA Amount	Ratio To Next Highest Contributor	$\log_{10}(LR)$
75:20:2:2:1 1.0 ng A *N*-2	1	78	1451	4.16:1	27.70
	2	19	349	7.12:1	28.95
	3	3	49	-	-30
	4	-	-	-	-30
	5	-	-	-	-30
75:20:2:2:1 1.0 ng A *N*-1	1	78	1410	4.43:1	27.45
	2	18	318	6.63:1	27.88
	3	3	48	1.71:1	0.74
	4	2	28	-	6.75
	5	-	-	-	-0.84
75:20:2:2:1 1.0 ng A *N*	1	78	1420	4.55:1	27.32
	2	17	312	6.37:1	27.70
	3	3	49	1.63:1	0.76
	4	2	30	3.00:1	5.46
	5	1	10	-	0.42
75:20:2:2:1 1.0 ng B *N*-2	1	81	1353	5.11:1	27.68
	2	16	265	5.52:1	28.77
	3	3	48	-	-30
	4	-	-	-	-30
	5	-	-	-	-30
75:20:2:2:1 1.0 ng B *N*-1	1	80	1388	5.38:1	27.49
	2	15	258	5.06:1	28.28
	3	3	51	1.82:1	2.08
	4	2	28	-	6.11
	5	-	-	-	-1.75
75:20:2:2:1 1.0 ng B *N*	1	81	1365	5.37:1	27.39
	2	15	254	5.77:1	28.10
	3	3	44	1.76:1	1.60
	4	1	25	3.13:1	5.53
	5	0	8	-	-0.51

four (*N*-1), and five (*N*) contributors. The *LR* for the two major known contributors were not affected significantly by varying the assigned *N* outside that expected from the MCMC variability. The *LR*s of the true minor contributors were affected when the assignment of *N* decreased. Some low level inclusionary *LR*s (maximum $LR = 35$) became exclusionary *LR*s ($LR < 1$) or exclusions ($LR = 0$, plotted as $\log_{10}(LR) = -30$ in Fig. 2) when *N* was underassigned relative to experimentally designed *N*.

It is somewhat intuitive that more obvious effects to the minor component *LR*s were observed for mixture ratios where more data was detected representing the minor components and where dropout is less likely. In addition, falsely constraining a mixture to a single minor contributor results in an increase in false exclusions (e.g. 48:48:2:2 at N-1 or 48:47:2:2:1 at N-2). Mixtures such as 48:48:2:2 (Fig. 3) demonstrated minor components that produced an *LR* in support of exclusion under *N*-1 and then an *LR* in support of inclusion under *N*.

The most informative series of mixtures examined with respect to discerning a difference between the perceived major components and the minor components were from the samples with mixture proportion 75:20:3:2 (Fig. 4) and 75:20:2:2:1 (Fig. 5). These were mixtures where the lowest major was closest to the highest minor and/or highest sum of the minors. The two major contributors to sample 75:20:3:2 have mixture proportions 0.75 (with $\log_{10}(LR) \approx 27$) and 0.20 (with $\log_{10}(LR) \approx 24$). Within Fig. 4 we see that the *LR*s for the two major contributors are not affected by an underestimation in *N* (*N*-1 compared with *N*). The APH for the contributor at proportion 0.20 was approximately 140 rfu for both replicates. The APH for the two minor contributors (mixture proportions 0.03 and 0.02) was 47 rfu (replicate A) and 64 rfu (replicate B). The ratio of the second major (proportion 0.20) to the sum of the minors was approximately 3:1 (replicate A) and 2:1 (replicate B) as determined by STRmix™. The *LR*s for the minor contributors were affected by varying *N* with some false exclusions occurring. A similar trend was seen for sample 75:20:2:2:1 (Fig. 5). The effect was a decrease in *LR* or a change to outright exclusion of a true contributor if an incorrect *N* was used to interpret the profile. This decrease in LR may be viewed as conservative in favour of the POI.

In Table 4, the template amount of DNA for each contributor to the replicate amplification of the 75:20:3:2 samples as determined by STRmix™ are given. The ratio of DNA from each major to the sum of the remaining lesser components is also given. The goal of this study was to attempt to determine a threshold or set of parameters that could be used to determine when a major component could be interpreted in the

Table 8
STRmix™ assigned values for mixture proportion, DNA amount, and the $\log_{10}(LR)$ along with the calculated ratio to the next highest component for each contributor to replicates of the 75:20:2:2:1 mixture (0.5 ng template DNA). Highlighted rows indicate components where the *LR*s were significantly affected by *N*. Exclusions ($LR = 0$) are represented as $\log_{10}(LR) = -30$.

Mixture Replicate *N*	Known Contributor	Mixture Proportion	DNA Amount	Ratio To Next Highest Contributor	$\log_{10}(LR)$
75:20:2:2:1:0.5 ng A *N*-2	1	78	779	4.53:1	27.58
	2	17	172	3.25:1	23.91
	3	5	53	-	-30
	4	0	0	-	-30
	5	0	0	-	-19.22
75:20:2:2:1:0.5 ng A *N*-1	1	76	779	4.50:1	27.36
	2	17	173	3.09:1	22.97
	3	5	56	2.33:1	-0.31
	4	2	24	-	3.04
	5	0	0	-	-1.20
75:20:2:2:1:0.5 ng A *N*	1	74	782	4.23:1	27.25
	2	18	185	3.25:1	22.37
	3	5	57	2.38:1	0.13
	4	2	24	3.43:1	2.07
	5	1	7	-	-0.54
75:20:2:2:1:0.5 ng B *N*-2	1	80	637	4.79:1	27.56
	2	17	133	6.05:1	26.85
	3	3	22	-	-1.46
	4	0	0	-	1.53
	5	0	0	-	-7.00
75:20:2:2:1:0.5 ng B *N*-1	1	80	647	4.72:1	27.43
	2	17	137	5.96:1	26.51
	3	3	23	3.29:1	0.01
	4	1	7	-	2.12
	5	0	0	-	-0.46
75:20:2:2:1:0.5 ng B *N*	1	80	664	4.99:1	27.29
	2	16	133	6.05:1	26.36
	3	3	22	2.00:1	-0.02
	4	1	11	5.50:1	1.88
	5	0	2	-	-0.38

presence of an unknown number of minor contributors. After inspection of Figs. 4 and 5 and Tables 4 and 5 we can see that there are instances when the ratio of the lowest major to the sum of the minor components may be less than one. Even in these mixtures, there is sufficient data for the software to consistently separate components with *LR*s and this is unaffected by varying *N*.

Review of Tables 4 and 5 in conjunction with Tables 6 through 9 indicate when the proportion of a major component is greater than 10%, the DNA amount for that component, as determined by STRmix™, is greater than 50 rfu, and the ratio of that component is at least 1.5:1 to the next highest component, the *LR* will not be significantly affected by uncertainty in *N*. For the purposes of this discussion, a significant change is considered to be a switch supporting one proposition to the opposing proposition or an exclusion to an *LR* close to 1.

A plot of mixture proportion given the experimental design versus the difference between mixture proportion assigned after STRmix™ interpretation and the experimental design proportion for all mixtures interpreted where the interpretation *N* equalled target *N* is given in Fig. 6. There is very good alignment across a range of mixture proportions. Some variability is observed due to STRmix™ assigning more or less template than experimental design and due to variability in pipetting when creating the mixtures.

4. Discussion

One complication when interpreting complex mixed DNA profiles is the assignment of *N*. The "true" *N* within a casework profile is always unknown and unknowable. A number of researchers have demonstrated that there is a risk of under-assigning *N* if relying on allele count alone. This risk increases with increasing *N* but can be mitigated by testing more loci or by testing more discriminating loci e.g. SE33 [11,23]. However, *N* can be overstated in the presence of a larger than expected stutter [18]. If an analyst is unable to reliably assign the *N*, then it may be prudent to report that the profile is unsuitable for further interpretation.

This study examined a range of mixtures in an effort to determine when the ambiguity in *N* would potentially affect the interpretation and subsequent assignment of an *LR* and what mixture components would be significantly affected. If one or more mixture components are significantly affected by *N* and there is ambiguity in the *N*, then these components are not suitable for comparison. Ideally, this decision would be made prior to a comparison to a reference sample. Based on the data in this study, a set of parameters has been determined that provide a minimum threshold to identify the components of a mixture that are suitable for comparison purposes when the *N* cannot be

Table 9
STRmix™ assigned values for mixture proportion, DNA amount, and the $\log_{10}(LR)$ along with the calculated ratio to the next highest component for each contributor to replicates of the 75:20:2:2:1 mixture (0.25 ng template DNA). Highlighted rows indicate components where the *LR*s were significantly affected by *N*. Exclusions ($LR = 0$) are represented as $\log_{10}(LR) = -30$.

Mixture Replicate N	Known Contributor	Mixture Proportion	DNA Amount	Ratio To Next Highest Contributor	Log_{10} (*LR*)
75:20:2:2:1 0.25 ng A *N*-2	1	74	249	4.61:1	27.27
	2	16	54	1.64:1	16.12
	3	10	33	-	-5.23
	4	0	0	-	-3.77
	5	0	0	-	-11.56
75:20:2:2:1 0.25 ng A *N*-1	1	72	253	5.16:1	27.14
	2	14	49	1.48:1	14.71
	3	9	33	1.83:1	-1.32
	4	5	18	-	-0.65
	5	0	0	-	-2.32
75:20:2:2:1 0.25 ng A *N*	1	72	260	5.91:1	27.06
	2	12	44	1.47:1	13.53
	3	8	30	1.50:1	-0.81
	4	6	20	2.00:1	-0.33
	5	3	10	-	-1.54
75:20:2:2:1 0.25 ng B *N*-2	1	77	273	5.57:1	27.44
	2	14	49	1.53:1	18.68
	3	9	32	-	-30
	4	0	0	-	-30
	5	0	0	-	-6.56
75:20:2:2:1 0.25 ng B *N*-1	1	74	274	5.83:1	27.30
	2	13	47	1.52:1	15.51
	3	8	31	1.82:1	-0.95
	4	4	17	-	-2.12
	5	0	0	-	-0.95
75:20:2:2:1 0.25 ng B *N*	1	73	271	6.16:1	27.20
	2	12	44	1.47:1	14.36
	3	8	30	1.58:1	-0.44
	4	5	19	2.38:1	-1.53
	5	2	8	-	-0.64

reasonably assigned. When a component of a mixture had a proportion of at least 10%, a ratio of at least 1.5:1 to the next highest component, and a DNA amount (as reported by STRmix™) of at least 50 rfu, the *LR* of the major component for a true contributor was not significantly affected by varying *N*. A significant change was considered to be a switch supporting one proposition to the opposing proposition or an exclusion to an *LR* close to 1. In some interpretations, the *LR*s of the major components decreased approximately one order of magnitude when interpreted assuming $N+1$ as expected. These parameters were used to examine all sets of mixtures under varying *N*. In each case, analysis of the major component under varying *N* provided support for inclusion for the true contributor. In some of the profiles interpreted, the sum of the mixture proportions of the minor components was greater than that of the lowest major component. Allele stacking or sharing between the different contributors did not have an effect on the ability to interpret the major component. The guidelines described above may be used to determine when a major component *LR* will not significantly change with varying *N*. It is important to note that this does not mean *LR*s calculated for contributors not meeting these criteria are unreliable. They are likely to change with varying *N* but are still reliable, reproducible and informative depending on the APH of the contributor. This distinction is discussed in detail by Taylor [22].

In the profiles investigated for this work, an increase in the assigned *N* over target *N* resulted in an increase in non-contributors producing low-level *LR*s clustered around an *LR* of 1 (Fig. 1, $N+1$). This effect has been observed previously [19]. Of interest, the three highest *LR*s for non-contributors were mixtures analysed assuming *N*-1. As expected, where there are more data detected representing the minor component (s) of mixtures, greater differentiation between the *LR*s for true and non-contributors was observed. Also as expected interpretation assuming fewer than target *N* (*N*-1 and *N*-2 interpretations) leads to false exclusions of known contributors. These false exclusions were always the minor contributor to a profile.

The contributor proportions determined by STRmix™ were similar to those from the experimental design. Differences can be attributed to uncertainty in the DNA quantitation measurements and the variability inherent to the PCR amplification and other laboratory processes [24]. Fig. 6 has demonstrated sufficient alignment that conclusions drawn from STRmix™ assigned template (and subsequent mixture proportion) is a suitable proxy for APH and assist in determining the suitability of the profile for further interpretation.

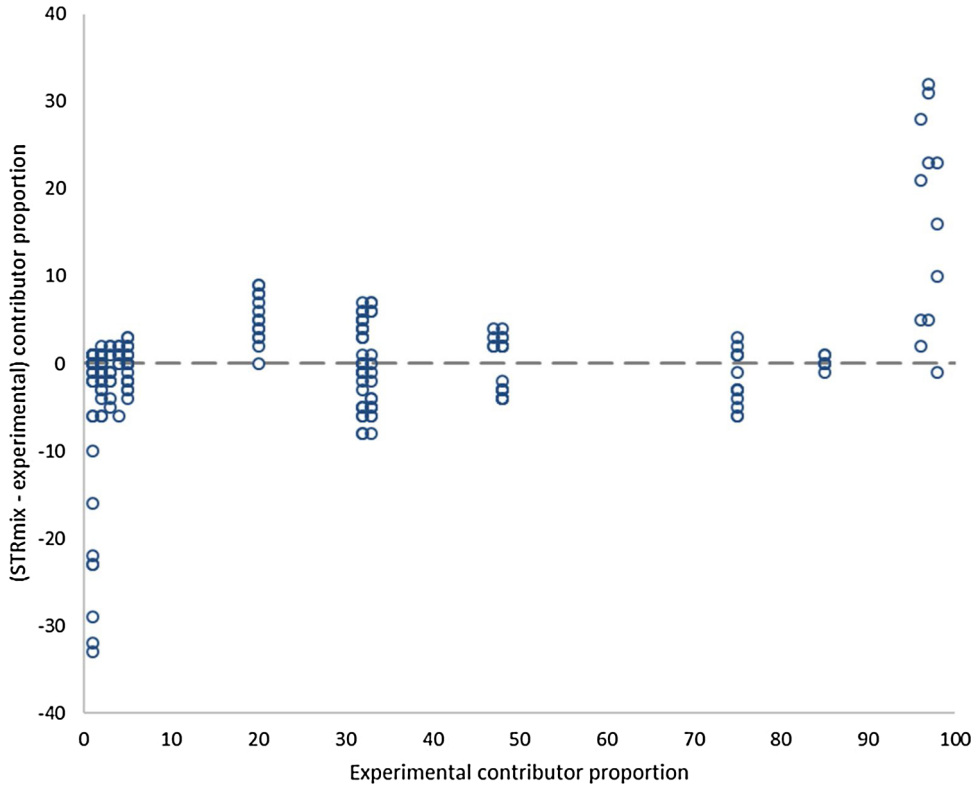

Fig. 6. Difference in the per contributor mixture proportion assigned by STRmix™ and the experimental design versus experimental design.

Acknowledgement

This work was supported in part by grant NIJ 2017-DN-BX-0136 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 organizations.

References

[1] P. Gill, L. Gusmão, H. Haned, W.R. Mayr, N. Morling, W. Parson, et al., DNA commission of the International Society of Forensic Genetics: recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods, Forensic Sci. Int. Genet. 6 (2012) 679–688.
[2] 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. 56 (2011).
[3] Ø Bleka, G. Storvik, P. Gill, 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. 21 (2016) 35–44.
[4] P. Gill, H. Haned, A new methodological framework to interpret complex DNA profiles using likelihood ratios, Forensic Sci. Int. Genet. 7 (2013) 251–263.
[5] D. Taylor, J.-A. Bright, J. Buckleton, The interpretation of single source and mixed DNA profiles, Forensic Sci. Int. Genet. 7 (2013) 516–528.
[6] Scientific Working Group on DNA Analysis Methods (SWGDAM). Guidelines for the Validation of Probabilistic Genotyping Systems, (2015) Accessed 3 October 2016 http://media.wix.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf.
[7] K. Slooten, A. Caliebe, Contributors are a nuisance (parameter) for DNA mixture evidence evaluation, Forensic Sci. Int. Genet. 37 (2018).
[8] J.S. Buckleton, J.-A. Bright, S. Gittelson, T.R. Moretti, A.J. Onorato, F.R. Bieber, et al., The probabilistic genotyping software STRmix: utility and evidence for its validity, J. Forensic Sci. (2018), https://doi.org/10.1111/1556-4029.13898.
[9] J.-A. Bright, R. Richards, M. Kruijver, H. Kelly, C. McGovern, A. Magee, et al., Internal validation of STRmix™ – a multi laboratory response to PCAST, Forensic Sci. Int. Genet. 34 (2018) 11–24.
[10] 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.
[11] 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.
[12] M.A. Marciano, J.D. Adelman, PACE: probabilistic assessment for contributor estimation– a machine learning-based assessment of the number of contributors in DNA mixtures, Forensic Sci. Int. Genet. 27 (2017) 82–91.
[13] H. Swaminathan, C.M. Grgicak, M. Medard, D.S. Lun, NOCIt: a computational method to infer the number of contributors to DNA samples analyzed by STR genotyping, Forensic Sci. Int. Genet. 16 (2015) 172–180.
[14] H. Haned, L. Pène, J.R. Lobry, A.B. Dufour, D. Pontier, Estimating the number of contributors to forensic DNA mixtures: does maximum likelihood perform better than maximum allele count? J. Forensic Sci. 56 (2011) 23–28.
[15] A. Biedermann, S. Bozza, K. Konis, F. Taroni, 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. 6 (2012) 689–696.
[16] H. Haned, L. Pène, F. Sauvage, D. Pontier, The predictive value of the maximum likelihood estimator of the number of contributors to a DNA mixture, Forensic Sci. Int. Genet. 5 (2011) 281–284.
[17] D. Taylor, J.-A. Bright, J. Buckleton, Interpreting forensic DNA profiling evidence without specifying the number of contributors, Forensic Sci. Int. Genet. 13 (2014) 269–280.
[18] 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.
[19] 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.
[20] 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.
[21] C.R. Hill, D.L. Duewer, M.C. Kline, M.D. Coble, J.M. Butler, U.S. Population data for 29 autosomal STR loci, Forensic Sci. Int. Genet. 7 (2013) e82-e3.
[22] D. Taylor, Using continuous DNA interpretation methods to revisit likelihood ratio behaviour, Forensic Sci. Int. Genet. 11 (2014) 144–153.
[23] J.M. Curran, J. Buckleton, Uncertainty in the number of contributors for the European standard set of loci, Forensic Sci. Int. Genet. 11 (2014) 205–206.
[24] J.-A. Bright, K.E. Stevenson, J.M. Curran, J.S. Buckleton, The variability in likelihood ratios due to different mechanisms, Forensic Sci. Int. Genet. 14 (2015) 187–190.