Internal Validation of STRmix™ V2.4

Palm Beach County

Sheriff's Office (PBSO) Laboratory

(Fusion™ 5C)

REVIEWED - attp.
Date: 5/30/17
Initials:
OK to use on casework

STRmix™ internal validation

This document describes the internal validation of STRmix™ V2.4 for Fusion™ 5C, at the Palm Beach County Sheriff's Office Crime Laboratory, (PBSO) West Palm Beach, Florida.

STRmix™ has previously been subjected to developmental validation. This involved, in part, the complete 'by hand' confirmation of the calculations behind the software. The results of the developmental validation are included in the STRmix™ User's Manual. In addition, a summary of the developmental validation is discussed in Taylor et al. [1]. A list of all papers describing the theory behind different aspects of STRmix™ is provided in Appendix 1 of this document.

Internal validation describes the activities PBSO has undertaken in-house before the implementation of STRmix™ into routine casework. This document follows elements of the internal validation section of the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems [2]. This included investigations into specificity and sensitivity mixture studies. The section where specific SWGDAM guidelines are discussed in this document is cross referenced in Appendix 2.

The results of all experiments related to the internal validation of STRmix™ at PBSO Laboratory are retained in the companion binder.

STRmix™ parameters

The parameters described in the document "Estimation of STRmix™ parameters for PBSO" were used for all internal validation checks presented in this report. All other run parameters have been optimized by the STRmix™ developers.

Section A: Single source profiles

Inspection of weights

This section covers the following standards

4.1.5 Single-source specimens

4.2.1.2 For Single-source specimens with high quality results, genotypes derived from non-probabilistic analyses of profiles above the stochastic threshold should be in complete concordance with the results of probabilistic methods.

This section demonstrates how the weights assigned to different genotype combinations are appropriate. The weights can be used as a diagnostic of the deconvolution process and should be intuitively correct, where the most supported genotypes have the highest weights.

A dilution series of a single source profile where the peak heights ranged from above the level where dropout is observed to below was constructed. Profiles were amplified using PowerPlex® Fusion 5C as per protocol. The amount of DNA amplified was as follows 0.5 ng, 0.25ng, 0.125 ng, 0.06 ng, and 0.03 ng. The profiles were analysed for input into the STRmix™ software utilizing GMIDX v 1.5 software with an analytical threshold of 125 RFU for the blue dye, 128 RFU for the green dye, 143 for the yellow dye, and 118 for the red dye. The profiles were interpreted in STRmix™, calculating the likelihood ratio (LR) for the known contributor. Likelihood ratios were calculated using the propositions:

H_p: The DNA originated from the person of interest

H_d: The DNA originated from an unknown individual

The LR was calculated for the known contributor using the laboratory's Caucasian, African American, and Hispanic allele frequencies and F_{ST} (θ) 0f 0.01 for the 99.0% 1-sided lower HPD. A plot of log(LR) versus input DNA amount (ng) is provided in Figure A1.

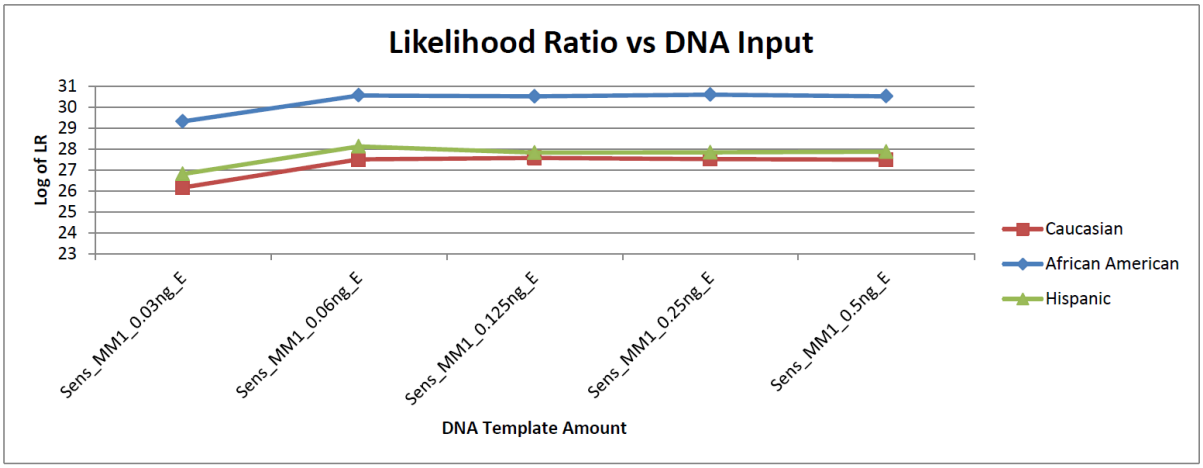


Figure A1. Plot of log(LR) versus input amount (ng).

Figure A1 shows the LR progressing from the value for the single source LR calculated for a full profile at >0.125 ng towards LR =1 as the DNA template decreases. As expected, the weights for genotypes considering dropout increased as DNA template drops. In addition, the DNA amounts from the STRmix™ output (*t* or template mass parameter) decline steadily in line with peak heights (data not shown).

Reproduction of Single Source LR

The total LR calculated by the STRmix™ for single source profiles can be easily replicated by hand. "Hand" calculations were conducted using a Microsoft Excel template provided by ESR. The total LR can be calculated by hand using the Balding and Nichols formula. For single source profiles:

$$\frac{2[\theta+(1-\theta)pi][\ \theta+(1-\theta)pj]}{(1+\theta)(1+2\theta)}$$ for heterozygote loci (equation 1)

$$\frac{[3\theta+(1-\theta)pi][2\theta+(1-\theta)pi]}{(1+\theta)(1+2\theta)}$$ for homozygote loci (equation 2)

Where p_i is the allele frequency for allele *i*, p_j the allele frequency for allele *j* and θ is the F_{ST} value of 0.01. The allele frequency used within equations 1 and 2 are posterior mean frequencies. These are calculated using the following equation:

$$\frac{\mathrm{x}i \; + \; \frac{1}{k}}{\mathrm{N}\alpha + 1} \qquad \text{(equation 3)}$$

where x_i is the number of observations of allele i in a database, N_a is the number of alleles in that database, and k is the number of allele designations with non-zero observations in the database.

Setting θ to zero returns the product rule where:

2_{pipj} for heterozygote loci (equation 4)

$_{pi}^{2}$ for homozygote loci (equation 5)

The 'by hand' calculated and STRmix™ results for the four single source profiles with a theta of 0.01 are listed in Tables A1-A5.

Table A1. Comparision of total likelihood ratios calculated by hand as compared to STRmix™.

LR Total

Samples compared to hand calculations

	Hand Calculations			STRmix		
Sample Name	LR (Caucasian)	LR (African American)	LR (Hispanic)	LR (Caucasian)	LR (African American)	LR (Hispanic)
Sens M1 0.5 ng	**6.77E+27**	**8.90E+30**	**3.01E+30**	**6.77E+27**	**8.90E+30**	**3.01E+28**
13-000071_4_6	**3.18E+28**	**1.05E+31**	**3.80E+29**	**3.18E+28**	**1.05E+31**	**3.80E+29**
14-000072_3_3	**1.75E+27**	**1.62E+30**	**3.83E+27**	**1.75E+27**	**1.62E+30**	**3.83E+27**
13-000074_3_4	**4.62E+27**	**1.40E+31**	**4.94E+28**	**4.25E+27**	**1.40E+31**	**4.94E+28**

Table A2. Comparison of likelihood ratios for each allele calculated by hand as compared to STRmix™ for sample M1.

M1 0.5ng Total LR

	Caucasian		African American		Hispanic	
	LR Hand	LR STRmix™	LR Hand	LR STRmix™	LR Hand	LR STRmix™
D3S1358	7.5	7.5	5.0	5.0	5.5	5.5
D1S1656	26.1	26.1	42.1	42.1	35.9	35.9
D2S441	33.5	33.5	102.5	102.0	27.7	27.7
D10S1248	5.4	5.4	7.5	7.5	5.3	5.3
D13S317	12.2	12.2	5.4	5.4	15.5	15.5
Penta_E	27.9	27.9	63.3	63.3	50.9	50.9
D16S539	5.0	5.0	7.6	7.6	6.7	6.7
D18S51	18.6	18.6	16.6	16.6	23.1	23.1
D2S1338	66.1	66.1	190.9	191.0	107.7	108.0
CSF1PO	4.5	4.5	6.7	6.7	4.7	4.7
Penta_D	11.0	11.0	31.6	31.6	13.6	13.6
TH01	19.9	19.9	7.4	7.4	11.0	11.0
vWA	21.5	21.5	16.4	16.4	17.9	17.9
D21S11	18.0	18.0	41.0	41.0	13.5	13.5
D7S820	81.9	81.9	244.8	245.0	152.4	152.0
D5S818	113.6	114.0	105.0	105.0	121.1	121.0
TPOX	3.8	3.8	6.2	6.2	4.0	4.0
D8S1179	26.0	26.0	18.7	18.7	27.1	27.1
D12S391	56.9	56.9	63.3	63.3	32.6	32.6
D19S433	30.7	30.7	33.3	33.3	22.2	22.2
FGA	24.3	24.3	45.6	45.6	42.4	42.4
D22S1045	6.3	6.3	22.4	22.4	7.5	7.5

Table A3. Comparison of likelihood ratios for each allele calculated by hand as compared to STRmix™ for sample 13-000071_4_6.

13-000071_4_6 Total LR

	Caucasian		African American		Hispanic	
	LR Hand	LR STRmix™	LR Hand	LR STRmix™	LR Hand	LR STRmix™
D3S1358	33.9	33.9	155.6	156.0	48.0	48.0
D1S1656	37.0	37.0	11.7	11.7	28.1	28.1
D2S441	6.7	6.8	14.9	14.9	4.9	4.9
D10S1248	67.1	67.1	72.3	72.3	66.4	66.4
D13S317	12.2	12.2	5.4	5.4	15.5	15.5
Penta_E	40.9	40.9	74.2	74.2	40.0	40.0
D16S539	13.9	13.9	12.7	12.7	12.2	12.2
D18S51	23.6	23.6	34.9	34.9	25.0	25.0
D2S1338	50.3	50.3	107.1	107.0	67.1	67.1
CSF1PO	7.1	7.1	10.2	10.2	6.6	6.6
Penta_D	15.8	15.8	59.2	59.2	29.7	29.7
TH01	7.7	7.7	71.0	71.0	17.8	17.8
vWA	15.5	15.5	10.6	10.6	13.3	13.3
D21S11	14.6	14.6	9.6	9.6	22.1	22.1
D7S820	12.8	12.8	6.4	6.4	12.7	12.7
D5S818	25.5	25.5	32.7	32.7	23.9	23.9
TPOX	16.4	16.4	14.5	14.5	18.0	18.0
D8S1179	34.3	34.3	59.7	59.7	59.8	59.8
D12S391	20.8	20.8	26.9	26.9	22.9	22.9
D19S433	5.3	5.4	9.3	9.3	6.2	6.2
FGA	16.9	16.9	17.6	17.6	19.8	19.8
D22S1045	100.404	100.0	81.0	81.0	72.4	72.4

Table A4. Comparison of likelihood ratios for each allele calculated by hand as compared to STRmix™ for sample 14-000072_3_3.

14-000072_3_3 Total LR

	Caucasian		African American		Hispanic	
	LR Hand	LR STRmix™	LR Hand	LR STRmix™	LR Hand	LR STRmix™
D3S1358	18.9	18.9	18.7	18.7	24.0	24.0
D1S1656	51.7	51.7	78.2	78.2	54.5	54.5
D2S441	9.5	9.5	20.0	20.0	7.0	7.0
D10S1248	21.4	21.4	21.2	21.2	18.7	18.7
D13S317	14.9	14.9	8.2	8.2	18.5	18.5
Penta_E	30.6	30.6	40.6	40.6	39.6	39.6
D16S539	9.1	9.1	19.9	19.9	11.5	11.5
D18S51	27.3	27.3	126.0	126.0	23.1	23.1
D2S1338	21.5	21.5	50.7	50.7	33.9	33.9
CSF1PO	6.2	6.2	6.6	6.6	5.5	5.5
Penta_D	20.3	20.3	51.4	51.4	20.5	20.5
TH01	10.5	10.5	8.9	8.9	6.9	6.9
vWA	11.8	11.8	12.7	12.7	9.4	9.4
D21S11	18.0	18.0	41.0	41.0	13.5	13.5
D7S820	15.9	15.9	10.4	10.4	14.0	14.0
D5S818	9.4	9.4	9.2	9.2	11.1	11.1
TPOX	3.5	3.5	6.8	6.9	4.0	4.0
D8S1179	13.8	13.8	56.1	56.1	18.0	18.0
D12S391	118.3	118.0	38.5	38.5	116.0	116.0
D19S433	11.9	11.9	22.7	22.7	15.5	15.5
FGA	18.3	18.3	39.2	39.2	31.7	31.7
D22S1045	69.8	69.8	31.3	31.3	52.4	52.4

Table A5. Comparison of likelihood ratios for each allele calculated by hand as compared to STRmix™ for sample 13-000074_3_4.

13-000074_3_4

	Caucasian		African American		Hispanic	
	LR Hand	LR STRmix™	LR Hand	LR STRmix™	LR Hand	LR STRmix™
D3S1358	8.4	8.4	7.4	7.4	8.2	8.2
D1S1656	37.0	37.0	11.7	11.7	28.1	28.1
D2S441	5.9	5.9	5.1	5.1	7.9	7.9
D10S1248	17.8	17.8	25.9	25.9	21.6	21.6
D13S317	27.4	27.4	40.1	40.1	21.0	21.0
Penta_E	101.8	101.8	107.6	107.6	61.7	61.7
D16S539	44.3	44.3	69.8	69.8	79.4	79.4
D18S51	23.2	23.2	35.5	35.5	22.8	22.8
D2S1338	25.2	25.2	49.4	49.4	44.4	44.4
CSF1PO	4.5	4.5	6.7	6.7	4.7	4.7
Penta_D	11.0	11.0	31.6	31.6	13.6	13.6
TH01	15.5	15.5	42.8	42.8	15.0	15.0
vWA	8.4	8.4	13.5	13.5	10.8	10.8
D21S11	852.2	852.0	1002.4	1002.4	578.1	578.1
D7S820	17.5	17.5	42.8	42.8	31.5	31.5
D5S818	3.6	3.6	5.7	5.7	3.8	3.8
TPOX	3.8	3.8	6.2	6.2	4.0	4.0
D8S1179	102.4	102.4	173.8	173.8	172.0	172.0
D12S391	20.8	20.8	26.9	26.9	22.9	22.9
D19S433	13.4	13.4	14.4	14.4	17.2	17.2
FGA	13.0	13.0	19.0	19.0	18.5	18.5
D22S1045	4.0	4.0	10.0	10.0	3.4	3.4

The results in Tables A1-A5 show that STRmix™ is giving the expected answer based on the population genetic model being used.

Section B: Use of peak heights

This section covers the following standard:

4.1.4 Allelic peak height, to include off- scale peaks

STRmix™ is a fully continuous model that uses peak heights to inform the genotype combinations of contributors to profiles. As template decreases dropout starts to be considered. As the weights for genotypes considering dropout increase, the weights for genotype combinations for the *true* contributors decrease and subsequently the LR decreases.

STRmix™ treats all peaks that are above the saturation threshold (calculated as 30000 rfu for the Palm Beach County Sheriff's Office laboratory's Applied Biosystems 3500xl data) qualitatively and not quantitatively. It is not recommended that saturated profiles be interpreted within the STRmix™ as a profile that exceeds the saturation threshold is likely to have higher stutter peaks heights than expected by STRmix™.

Three single source samples with input amounts of DNA greater than 1 ng (containing rfu above 30000) were interpreted by STRmix™ to evaluate the impact of oversaturated data on profile interpretation and the weights assigned. The weights generated by STRmix™ were reviewed. All profiles were interpreted correctly, with weights =1 for the known genotype combination.

Section C: Weights

This section covers the following standard:

4.2.1.3 Generally, as the analyst's ability to deconvolute a complex mixture decreases so do the weightings of individual genotypes within a set determined by the software.

The weights are described as the primary output from STRmix™. They can be used as a diagnostic of the deconvolution process and should be intuitively correct, where the most supported genotypes have the highest weights. The ability to clearly define or identify the specific genotypes for a contributor to a mixture is reflected in the weights given to the genotype combinations. As the mixture proportions become more ambiguous (move toward 1:1 ratio), the weights of the genotype combinations for each contributor decrease.

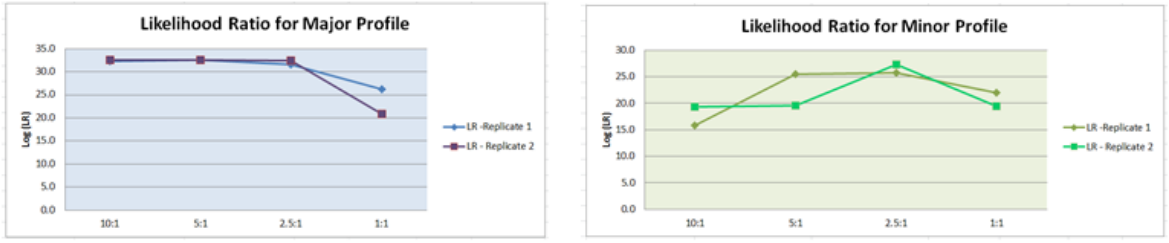
Two different two person mixture series in the following ratios 10:1, 5:1, 2.5:1 and 1:1 with a target DNA input of 0.5 ng were amplified in replicate and interpreted by STRmix™. The profiles were interpreted in STRmix™ under the following propositions and a LR calculated for the Caucasian, African American, and Hispanic sub populations:

H_p: The DNA originated from the person if interest (known major or minor) and an unknown individual

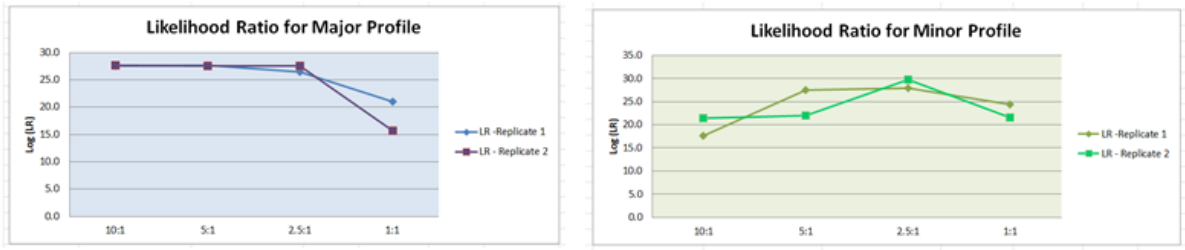
$H_{d:}$ The DNA originated from two unknown individuals

A plot of the log(LR) for the mixture series considering both the major and minor contributor for the three subpopulations is provided in Figures C1 and C2.

F2M2 African American



Caucasian



Hispanic

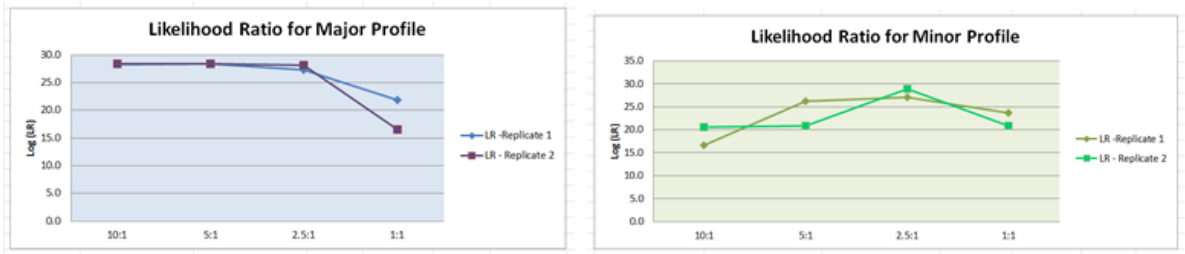
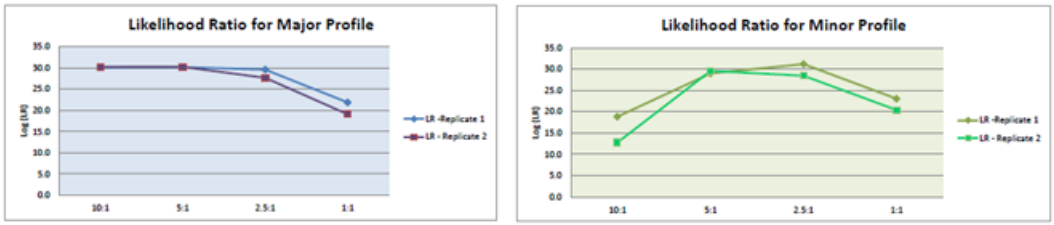
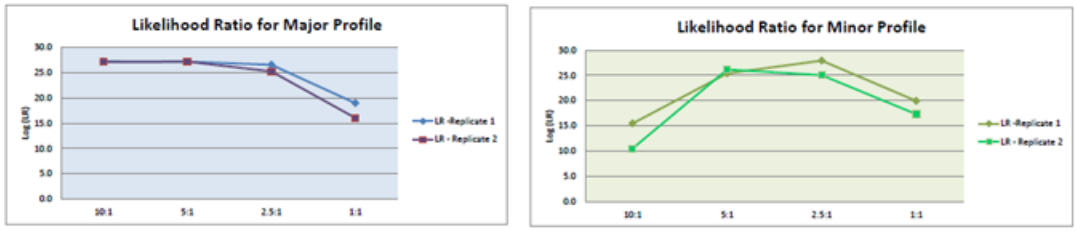


Figure C1. Log (LR) mixture ratio (proportion) for mixture set F2M5 considering both the major and minor contributor for the African American, Caucasian, and Hispanic subpopulation.

M1F1 African American



Caucasian



Hispanic

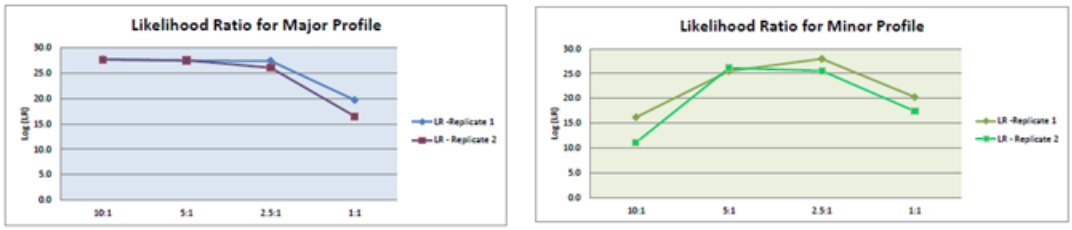


Figure C2. Log (LR) mixture ratio (proportion) for mixture set F2M5 considering both the major and minor contributor for the African American, Caucasian, and Hispanic subpopulation.

Figures C1 and C2 illustrate that the mixture proportions in the STRmix™ output changed appropriately as the mixture ratios varied. The log(LR) decreased by approximately half for the 1:1 mixture ratios when compared to the single source LR calculated for the major and minor contributors. The data presented in Figures C1 and C2 show that where the major contributor is in high template (500 pg) and has a high mixture proportion, the LR obtained approaches that obtained from the contributors single source profiles as the genotypes of the major contributor can be clearly resolved. When the mixture proportions move towards a 1:1 mixture ratio there are more genotype combinations possible to explain the observed profile and therefore the weights for the known major contributor's genotype combinations decrease as do the LR obtained. For the minor contributor, at low mixture proportions, the template amount of DNA for this contributor is low and some alleles may not be distinguished above the laboratory's analytical threshold or considered as possible stutter. An increase in mixture proportions and template DNA may allow more alleles to be both detected and distinguished from the major contributor thus causing the LR to rise. At mixture proportions close to 1:1 more alleles may be detected for the minor contributor, however, the LR may decrease due to ambiguity in assigning genotypes to individual contributors.

Section D: Sensitivity, specificity, and mixtures

This section covers the following standards:

4.1.2. Hypothesis testing with contributors and non-contributors

4.1.6. Mixed specimens

4.1.6.1. Various contributor ratios (e.g., 1:1 through 1:20, 2:2:1, 4:2:1, 3:1:1, etc.)

4.1.6.2. Various total DNA template quantities

4.1.6.3. Various numbers of contributors. The number of contributors evaluated should be based on the laboratory's intended use of the software. A range of contributor numbers should be evaluated in order to define the limitations of the software.

4.1.6.5. Sharing of alleles among contributors

4.1.7. Partial profiles, to include the following:

4.1.7.1. Allele and locus drop-out

4.1.13. Sensitivity, specificity, and precision as described for Developmental Validation

A demonstration of sensitivity and specificity for a range of PBSO Fusion™ 5C mixtures was undertaken as per Taylor [6]. With respect to interpretation methods, sensitivity is defined as the ability of the software to reliably resolve the DNA profile of known contributors within a mixed DNA profile for a range of starting DNA templates. The log(*LR*) for known contributors (H_p true) should be high and should trend to zero as less information is present within the profile. Information includes amount of DNA from the contributor of interest, conditioning profiles (for example the victim's profile on intimate samples), replicates and decreasing numbers of contributors. Specificity is defined as the ability of the software to reliably exclude known non contributors (H_d true) within a mixed DNA profile for a range of starting DNA templates. The log(*LR*) should trend upwards to zero as less information is present within the profile.

Specificity and sensitivity were tested by calculating the *LR* for a number of two, three, and four person profiles for both known contributors and known non-contributors. The plots in [6] have been reproduced for PBSO's Fusion® 5C data. Two sets of samples were generated for the two, three, and four contributor mixtures. One set contained profiles from mixtures of unrelated individuals and the second from mixtures of related individuals. A summary of the profiles is given in Table D1.

Table D1: Summary of experimental design for specificity and sensitivity tests.

Four person mixtures	**Three person mixtures**	**Two person mixtures**	**DNA amount of smallest contributor (pg)**	**Total target template (pg)**
4:4:1:1, 1:1:3:6 and 1:3:3:13*	1:1:8, 6:3:1, 5:5:1 and 1:3:3	N/A	62.5	1000
4:4:1:1, 1:1:3:6 and 1:3:3:13*	1:1:8, 6:3:1, 5:5:1 and 1:3:3	19:1, 1:19, 10:1, 5:1, 2.5:1, 1:2.5, 1:5 and 1:10	25	500
4:4:1:1, 1:1:3:6	1:1:8, 6:3:1, 5:5:1	19:1, 1:19, 10:1, 5:1,	12.5	250

and 1:3:3:13*	and 1:3:3	2.5:1, 1:2.5, 1:5 and 1:10		
4:4:1:1, 1:1:3:6 and 1:3:3:13*	1:1:8, 6:3:1, 5:5:1 and 1:3:3	10:1, 5:1, 2.5:1, 1:2.5, 1:5 and 1:10	6.25	100

*for 4 person mixtures at 1:3:3:13 the total template amounts were approximately 124.5, 310, 621.3 and 1242.5pg. These samples were labelled as 1:3:3:9

Four person mixtures	**Three person mixtures**	**Two person mixtures**	**Total target DNA template**
1:1:1:1 (25pg)	1:1:1 (33pg)	1:1 (50pg)	100
1:1:1:1 (62.5pg)	1:1:1 (82.5pg)	1:1 (125pg)	250
1:1:1:1 (125pg)	1:1:1 (165pg)	1:1 (250ug)	500
1:1:1:1 (250pg)	1:1:1 (333pg)	1:1 N/A	1000

The values in (brackets) in the above table are the approximate per contributor DNA amounts.

These profiles represent a spread of profile quality, including the 'worst' types of profiles likely to be encountered by the laboratory in casework. The profiles are of varying DNA quantity and mixture proportions. The contributors include homozygote and heterozygote alleles and there is varying amounts of allele sharing across the different loci (standard 4.1.6.5). Given the template amounts allele and/or locus dropout was expected to occur within the profiles containing the lower DNA amounts (standard 4.1.7.1).

Each profile was interpreted in STRmix™ V2.4.06 and compared to the known contributors for that particular mixture (H_p trues), the contributors to all the other mixtures provide by PBSO (H_d trues) [See Table D2] and 500 known non-contributors (H_d trues) using the Database Search function within STRmix™. The 500 non-contributors profiles were artificially generated using the NIST Caucasian Fusion® 5C allele frequency database.

Table D2: The known contributors to each of the mixture sets.

Mixture Set	**Contributor 1**	**Contributor 2**	**Contributor 3**	**Contributor 4**
2 person Unrelated	F2	M5	-	-
2 person Unrelated (19:1 mixtures)	PBSO F1	PBSO M1	-	-
2 person Unrelated (19:1 mixtures)	PBSO F2	PBSO M2	-	-
2 person Related	F1	M4	-	-
3 person Unrelated	M6	M7	F3	-
3 person Related	M8	M9	F4	-
4 person Unrelated	M10	M11	F5	F6
4 person Related	M1	M2	M3	F7

An LR was calculated using the Database Search function within STRmix™ considering the following propositions:

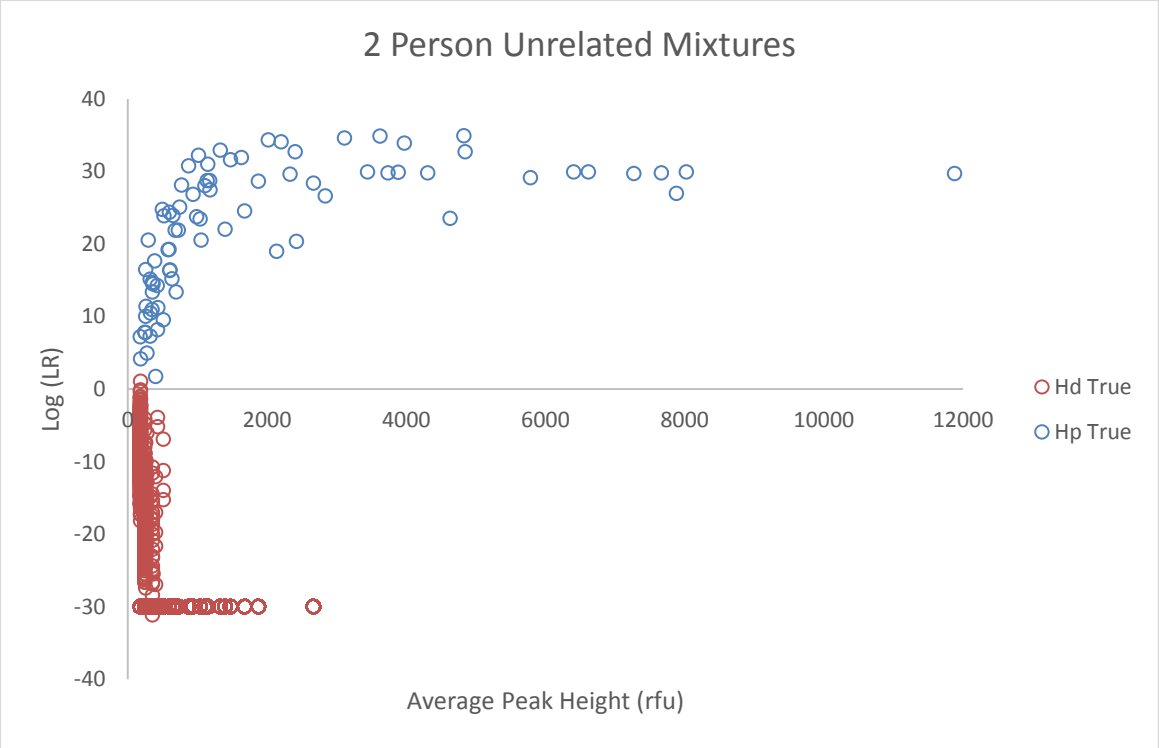
H_p: The DNA originated from the database individual and *N*-1 unknown individuals

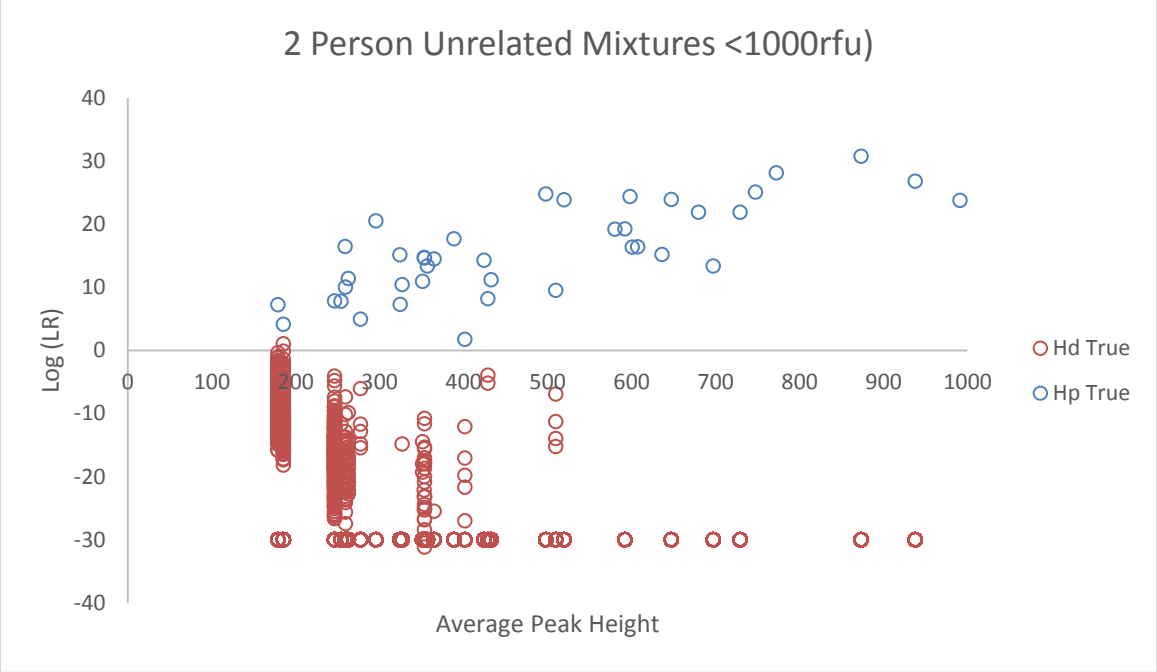
H_d: The DNA originated from *N* unknown individuals

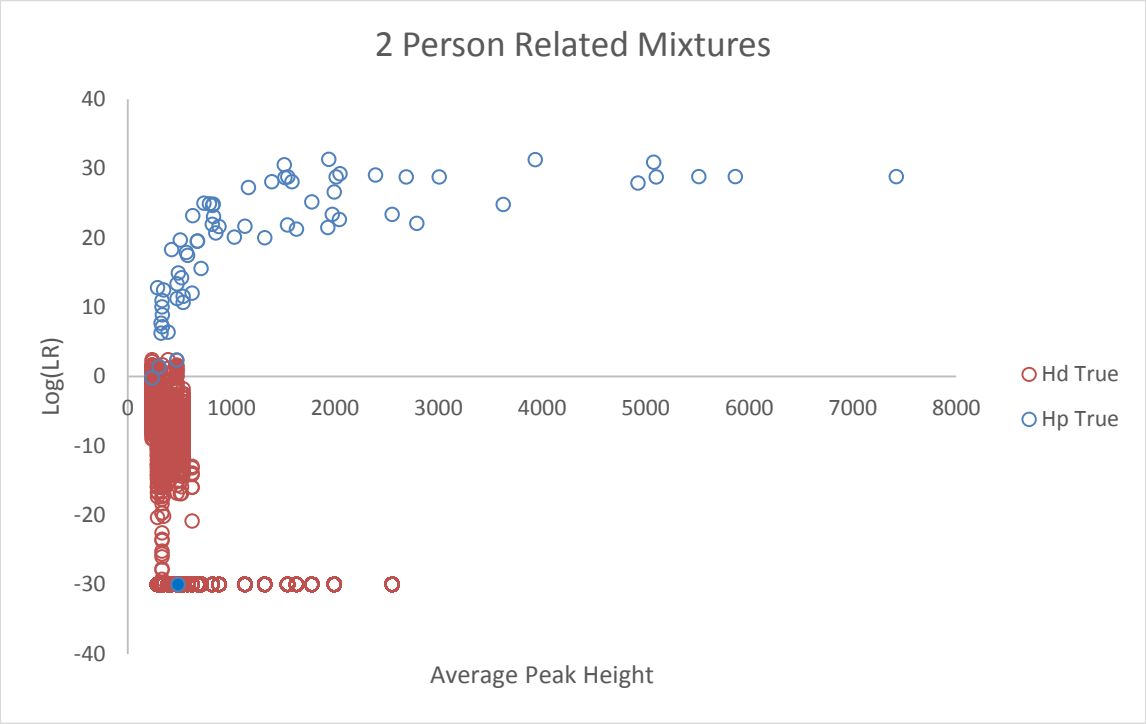
Where *N* is the number of contributors assumed in the mixture deconvolution.

An LR generated using this function is a product rule point estimate, without incorporation of F_{ST} or sampling uncertainty. The NIST (Fusion® 5C) Caucasian allele frequencies were used in these calculations.

Plots of log(*LR*) versus the average peak height (*APH*) per contributor for the two, three, and four contributor mixtures are given in Figure D1. Exclusions (*LR*=0) are plotted as log(*LR*)=-30. The *APH* per known contributor is taken from the unmasked and unshared alleles. The lowest contributor *APH* for each profile was used for the H_d true contributors. Where the *APH* for a known contributor was so low that it was unable to be calculated, a value of half (59 rfu) the lowest analytical threshold (118rfu) was used. The results of all comparisons are provided in Figure D1. It was not possible to determine the *APH* for all known contributors to the three person related mixture set due to masking and sharing of alleles. Plots of log(LR) versus amount of DNA (pg) or template are provided for this mixture set. The per contributor amount of DNA for H_d true (non)contributors is taken as the lowest of that for the known contributors.




2 Person Unrelated Mixtures <1000rfu)
Log (LR)
Average Peak Height
Hd True
Hp True


2 Person Related Mixtures
Log(LR)
Average Peak Height
Hd True
Hp True

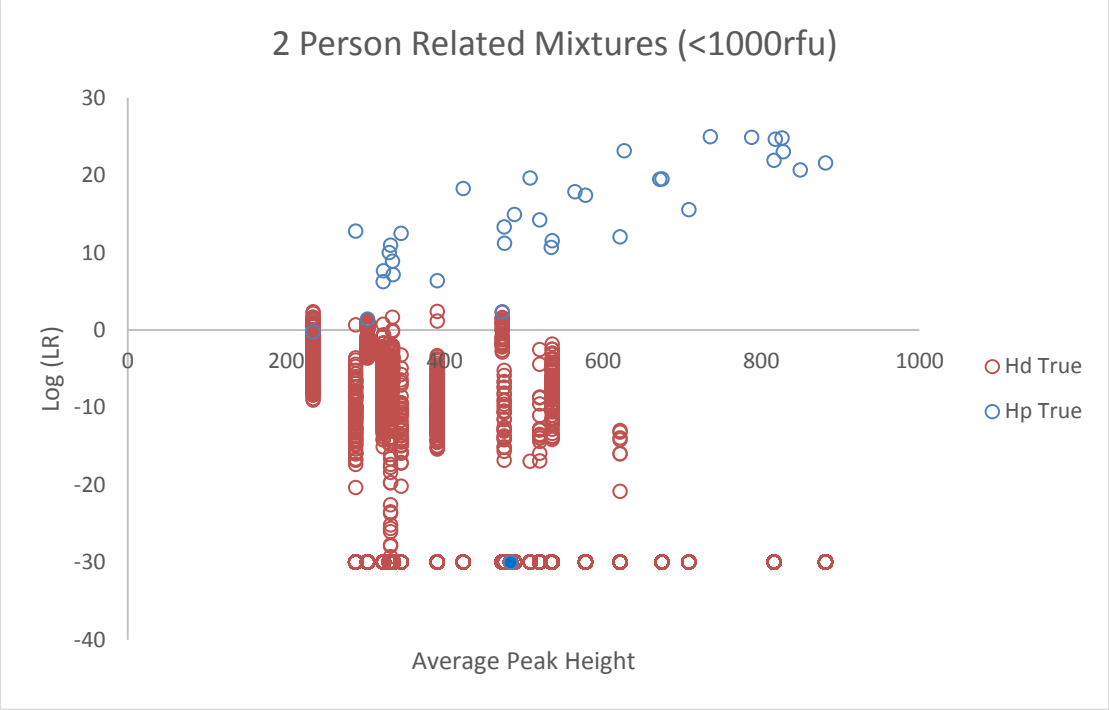
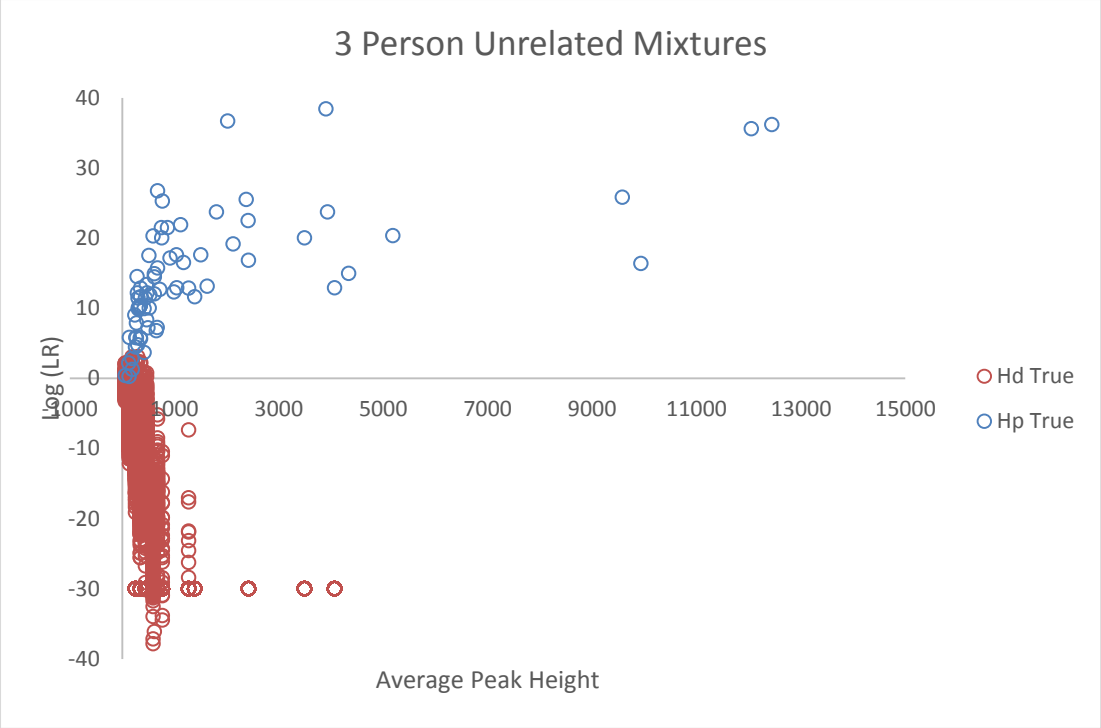
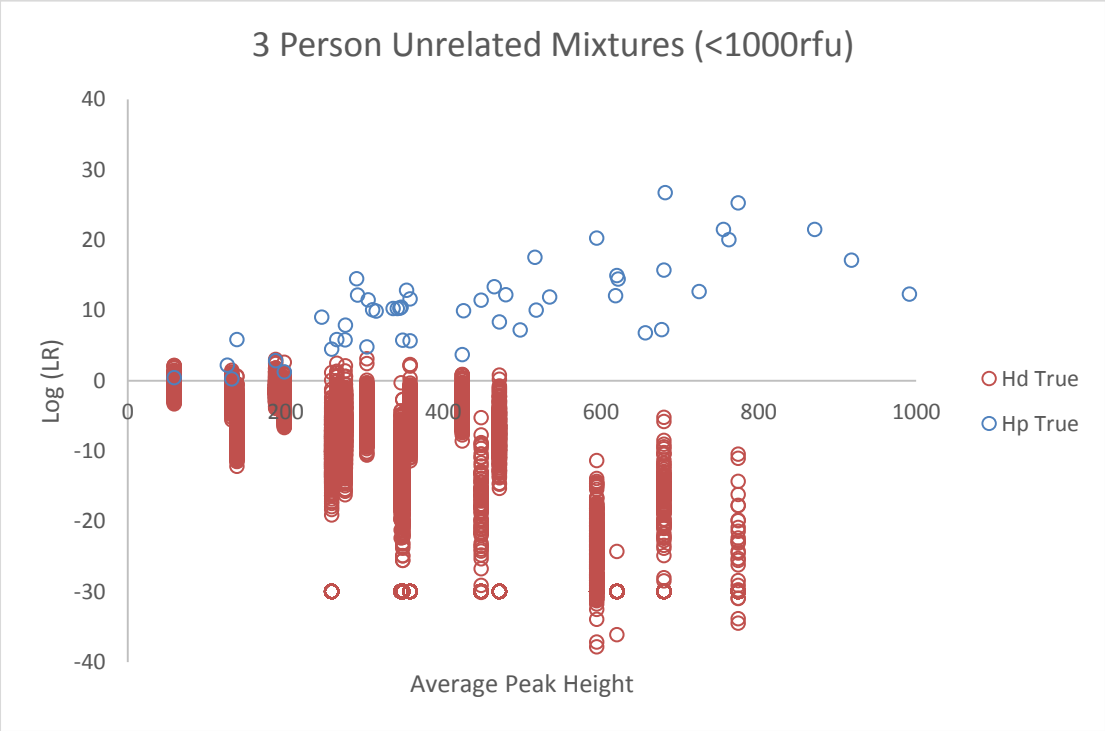


Figure D1: Log(LR) versus average peak height (APH in rfu) versus template (pg) for 2-, 3- and 4-person mixtures amplified using Fusion® 5C by the PBSO laboratory. The order of the plots goes Log(LR) vs APH (rfu), or template (pg), then a zoom of Log(LR) vs APH (rfu) or template(pg), for each 2-, 3- & 4-person mixture sets.

One data point has a -30 Log(*LR*) for M4 (known minor contributor). There is a small peak present at the '10' allele position at Penta D that could be interpreted as either possible 'drop-in' or truly allelic. If this peak is considered as 'drop-in', then positive Log(LR) values are obtained for M4 (9,14). If the '10' peak is considered as allelic, then the "9,14'' option for contributor two (minor) is excluded. This is a run to run variability issue. Reviewing the primary diagnostics (i.e., mixture proportions, weights, LR per locus, log likelihood, allele and stutter variance, Gelman Rubin Score, etc.) in the advance report may help to troubleshoot the exclusionary LR returned. In this instance the individual locus LR's, would help to identify this rare occurrence.


3 Person Unrelated Mixtures
40
30
20
10
0
-10
-20
-30
-40
Log (LR)
-1000
1000
3000
5000
7000
9000
11000
13000
15000
Average Peak Height
Hd True
Hp True


3 Person Unrelated Mixtures (<1000rfu)
40
30
20
10
0
-10
-20
-30
-40
Log (LR)
0
200
400
600
800
1000
Average Peak Height
Hd True
Hp True

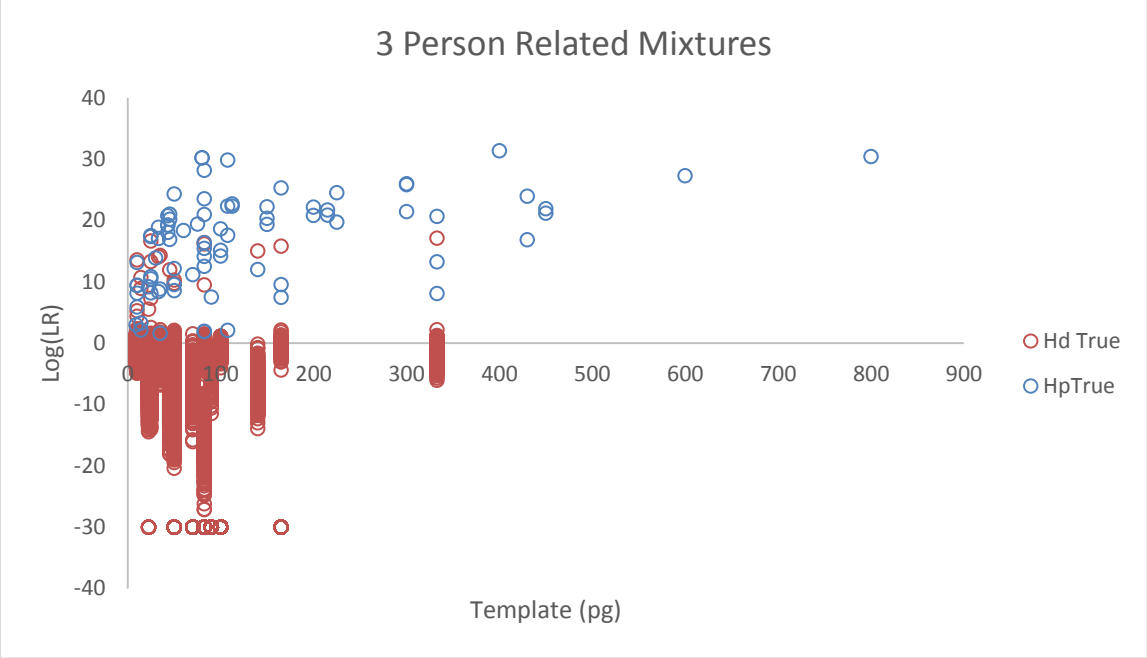
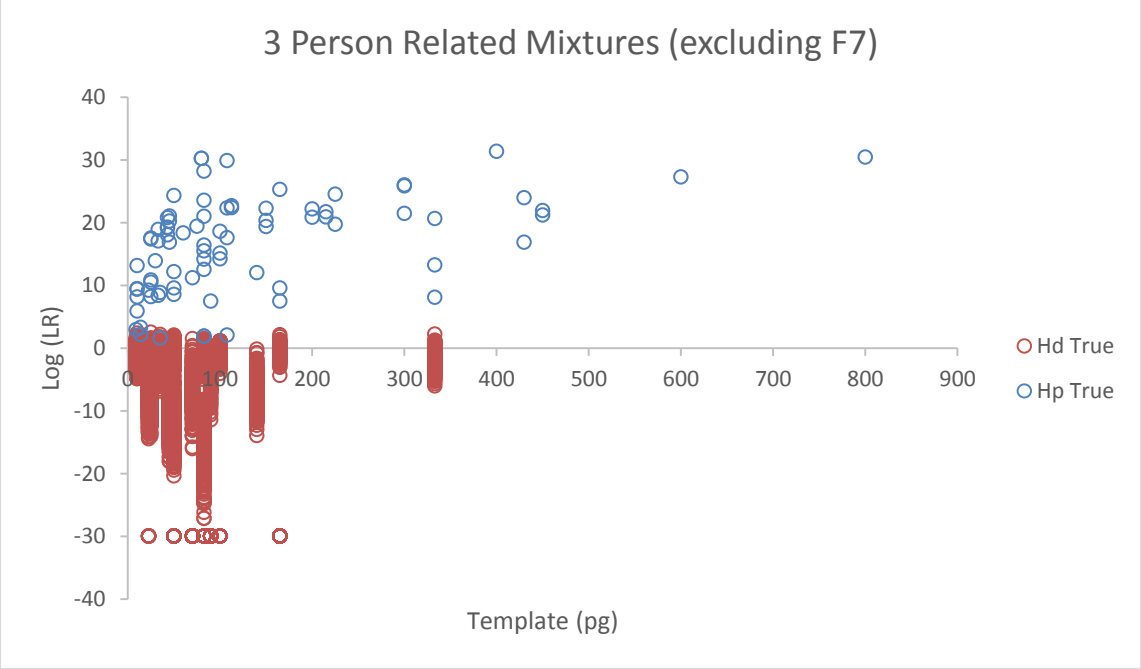
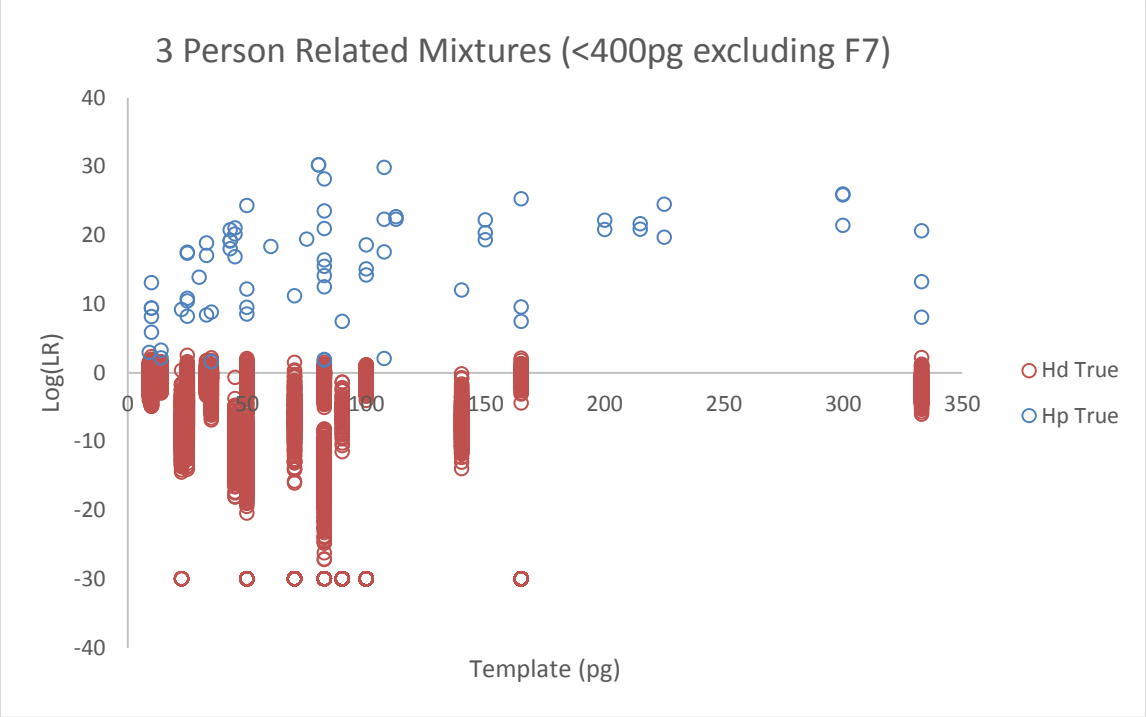
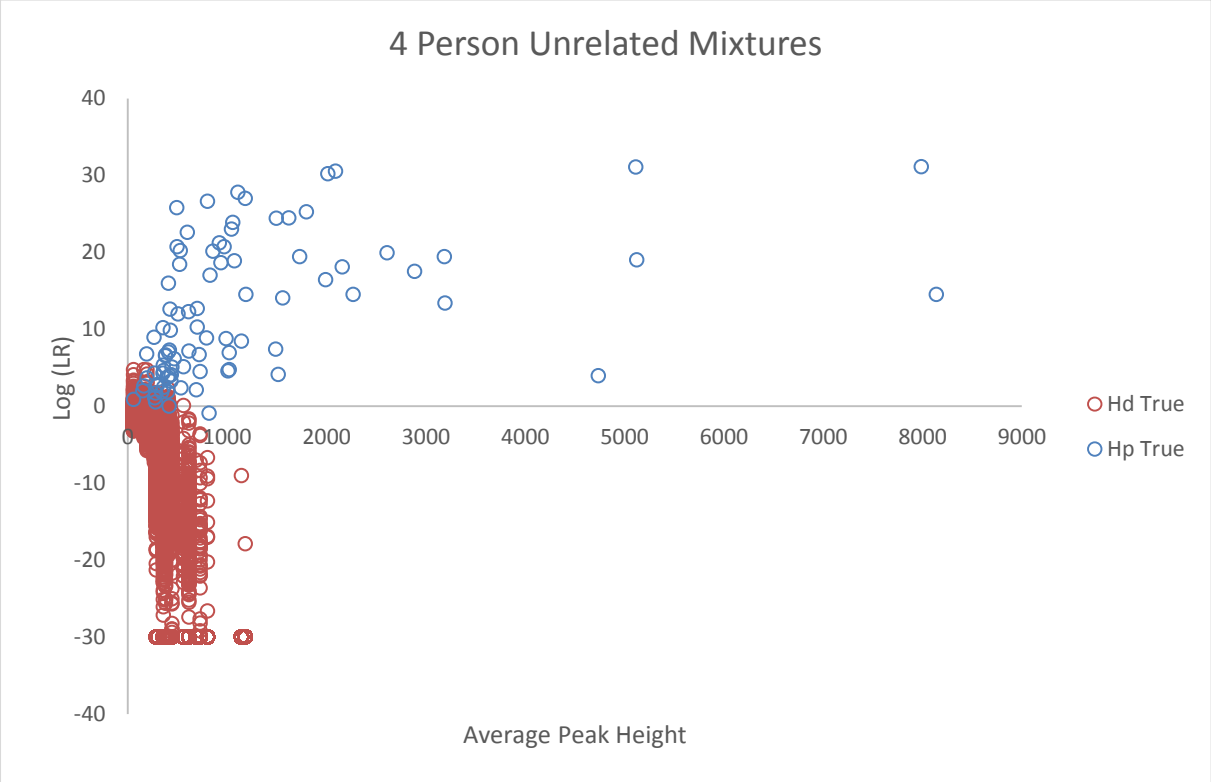


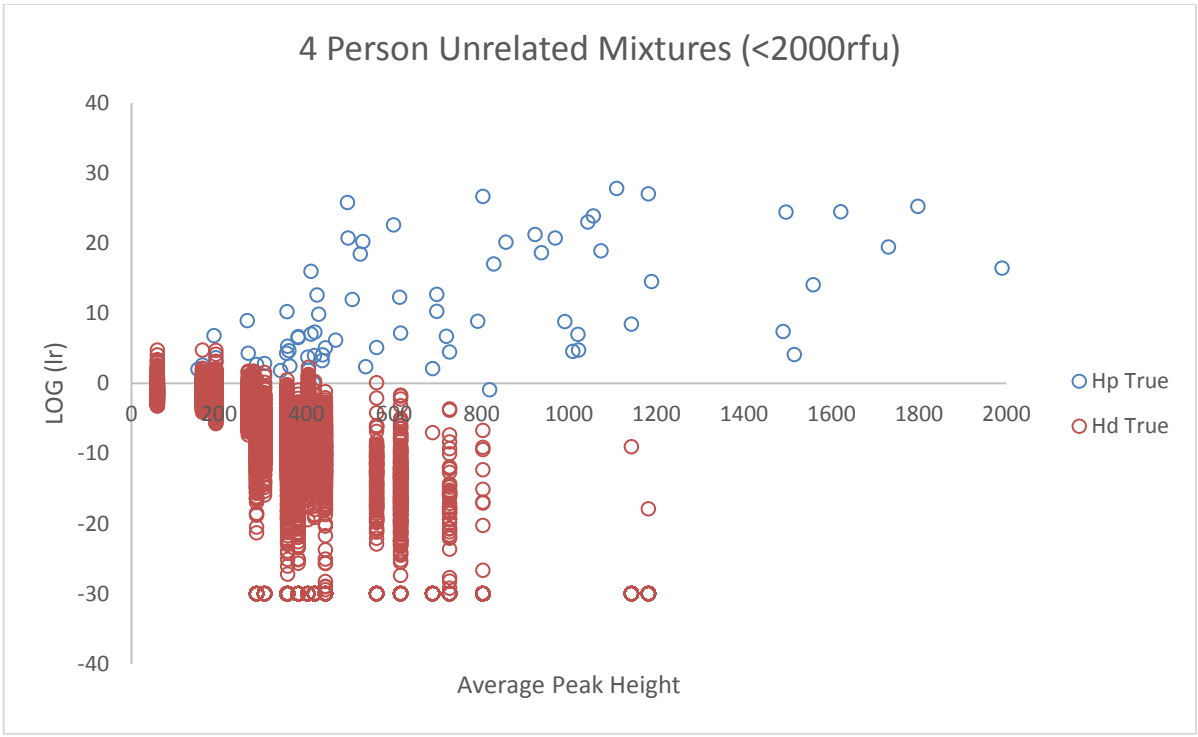
Figure D1: (continued): Log(LR) versus average peak height (APH in rfu) versus template (pg) for 2-, 3- and 4- person mixtures amplified using Fusion® 5C by the PBSO laboratory. The order of the plots goes Log(LR) vs APH (rfu), or template (pg), then a zoom of Log(LR) vs APH (rfu) or template(pg), for each 2-, 3- & 4-person mixture sets.

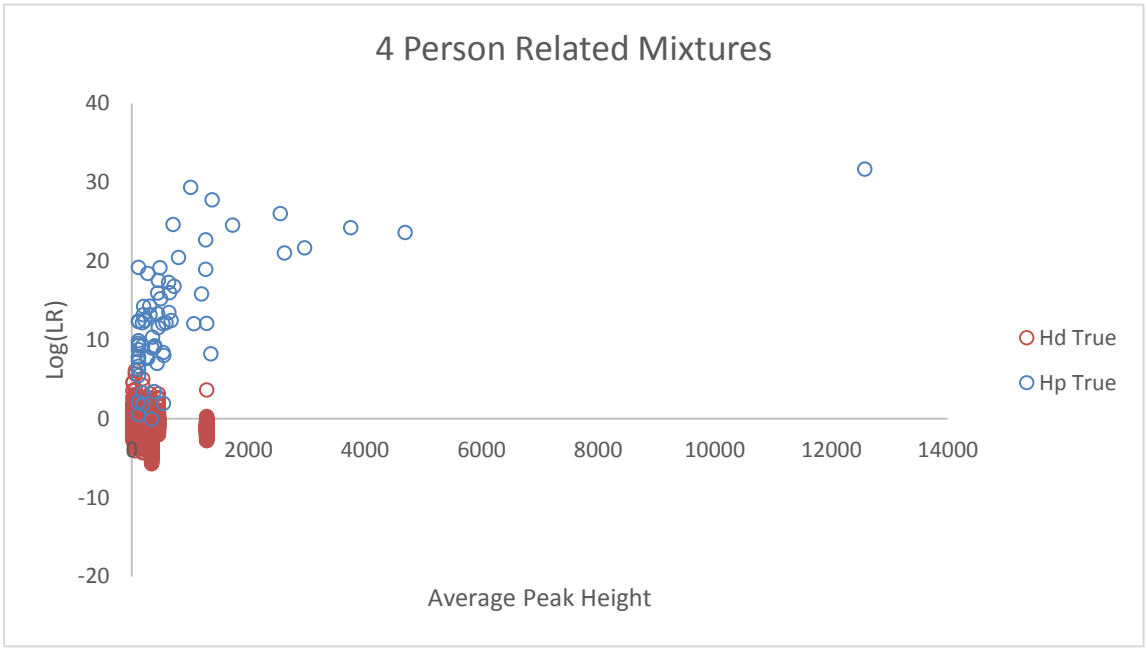
Positive Log(LR) values were obtained for known non contributor F7 in this data set. Inspection of F7's DNA profile indicated that she is a relative of M8, M9 and F4 and the data was replotted excluding the results for F7.




3 Person Related Mixtures (<400pg excluding F7)
Log(LR)
Template (pg)
Hd True
Hp True


4 Person Unrelated Mixtures
Log (LR)
Average Peak Height
Hd True
Hp True


4 Person Unrelated Mixtures (<2000rfu)
LOG (lr)
40
30
20
10
0
-10
-20
-30
-40
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Average Peak Height
Hp True
Hd True


4 Person Related Mixtures
Log(LR)
40
30
20
10
0
-10
-20
0
2000
4000
6000
8000
10000
12000
14000
Average Peak Height
Hd True
Hp True

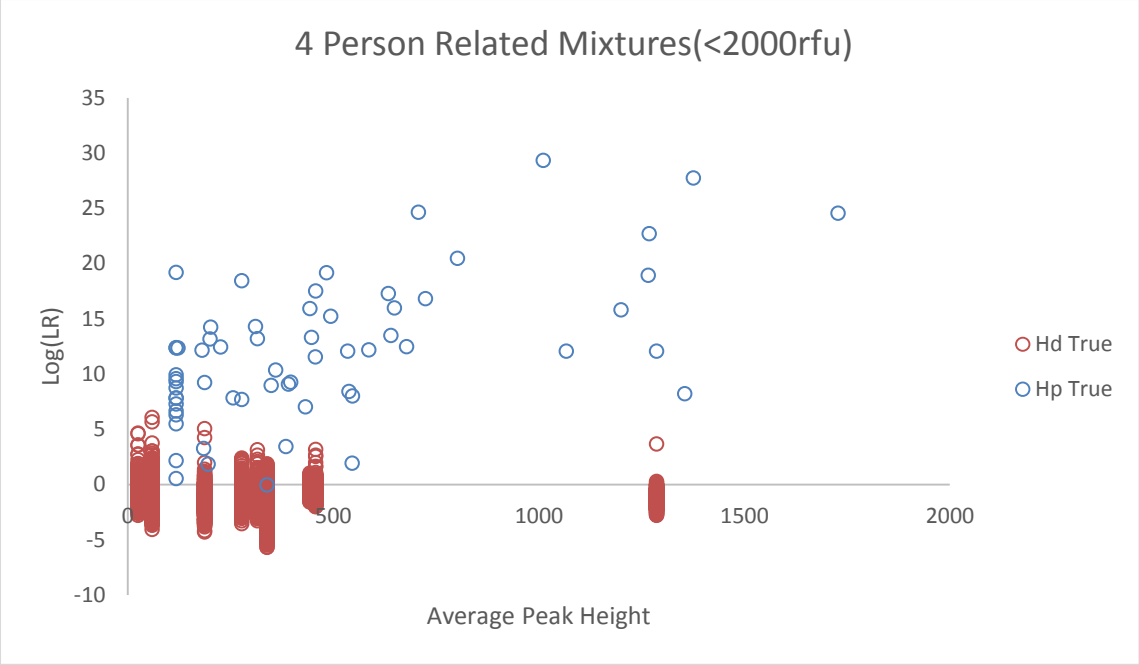


Figure D1: (continued): Log(LR) versus average peak height (APH in rfu) versus template (pg) for 2-, 3- and 4- person mixtures amplified using Fusion™ 5C by the PBSO laboratory. The order of the plots goes Log(LR) vs APH (rfu), or template (pg), then a zoom of Log(LR) vs APH (rfu) or template(pg), for each 2-, 3- & 4-person mixture sets.

Inspection of the plots in Figure D1 show the *LR* distributions for H_p true and H_d true were typically well separated at higher APH/template for 2-, 3- & 4-person mixtures. As the number of contributors increased and the template lowered, all the distributions converged on an *LR* of 1 [log(*LR*) = 0]. At high APH/template, STRmix™ correctly and reliably gave a high *LR* for true contributors and a low (or exclusionary) *LR* for false contributors. At low template or high contributor number STRmix™ correctly and reliably reported that the analysis of the sample tends towards uninformative or inconclusive.

The lowest H_p true *LR* was around 10^{-1} which was obtained in the 4-person unrelated mixture data and the highest H_d true *LR* was around 10^{+6} obtained in the 4-person related mixture set (excluding values for F7 from the 3 person related samples).The 3-person related mixture is comprised of two related donors and one unrelated donor. The 4-person related mixture is comprised of three related donors and one unrelated donor. The highest H_d true *LR* for the unrelated data set was 10^{+4}. It is also worth noting for all the unrelated mixtures no LRs greater than 1 were obtained when the APH is 600 rfu. With any statistical analysis there is the opportunity for adventitious matches to occur, where simply by chance a random individual could have similar DNA profiling results to the true contributor.

The plots in Figure D1 can help inform the limits of STRmix™, particularly the lower limit of DNA where an H_p true hypothesis results in a *LR* greater than 1 and the limit where false positives may arise (a *LR* greater than 1 where H_d is true).

Note. It is likely that the high allele and stutter variance values for this data (generated via the earlier Model Maker work up) will reduce the differentiation between true contributors and non-contributors. High variance values will make the acceptance rejection criteria for proposed genotype combinations during MCMC more tolerant of differences between observed and expected profiles (accepting more combinations) but close matches between the observed and the expected will be less well rewarded than they would be under lower variance values. Therefore there is likely to be a wider zone of LRs that might be considered "uninformative" and this 'zone' may come into play at higher values of template/APH than with platforms with lower variance.

Where profile peak heights or templates are low and for four person mixtures it may be beneficial to run replicate amplifications and/or increased iterations.

Additional review of mixtures:

A STRmix™ output contains run diagnostics to assist the user and to give confidence that the analysis has run as expected. Diagnostic values for all the mixtures described within section D of this report were collated. A summary of this information is provided in the following plots.

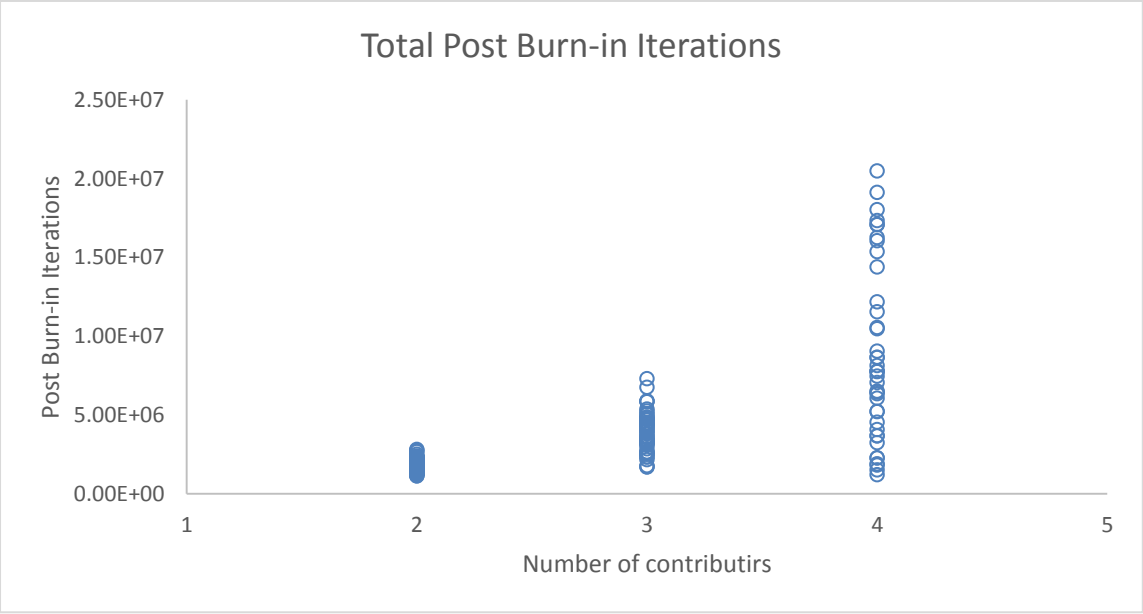


Figure D2: Plot of total iterations required to achieve 400,000 post burn-in accepts versus the true number of contributors to the mixtures.

Inspection of Figure D2 shows the expected increase in the number of iterations required for higher order mixtures. The more complex the mixture the greater the amount of iterations required to achieve the number of accepts, this can be orders of magnitudes more from 2 to 4 person mixtures.

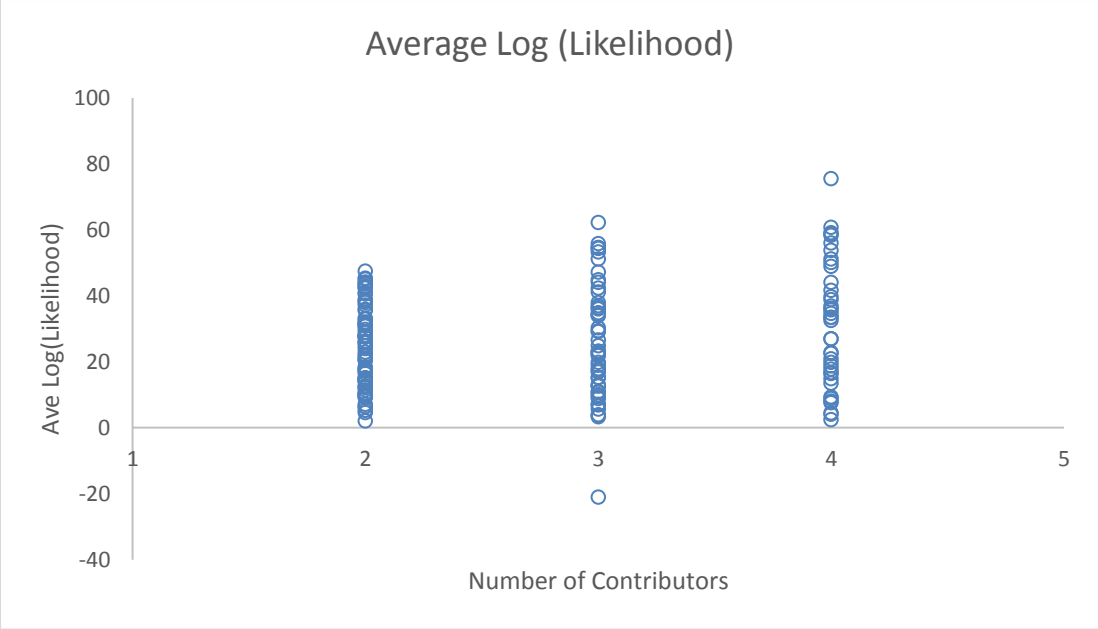


Figure D3: Plot of average Log(Likelihood) from the post burn-in compared to the true number of contributors.

Inspection of Figure D3 shows a spread of average Log(Likelihood) or probability density values ranging from a Log(Likelihood) of around 2 to over 70. It is anticipated that these values will range and broadly speaking the higher the better. However, low or even negative numbers do not preclude the use of the results.

There was one significant outlier with Log(Likelihood) of -21.08. This was from sample F07 5-5-1 1.0 ng within the 3 person unrelated set. This deconvolution also had a very high stutter variance at 505.3. This profile had increased baseline noise, affecting calling (or not) of stutter peaks, notably at D16, a D16 '13' allele had a peak height greater than saturation (30000 rfu). There was also a D12 '18' split peak, with no stutter resolved for large 19.1 allele.

However, low or negative Log(Likelihood) values alone do not necessarily invalidate results. These diagnostics should be considered as a whole, together with the users review of the primary diagnostics; the weights, proposed mixture proportions (Mx) and (if applicable) the individual locus LRs.

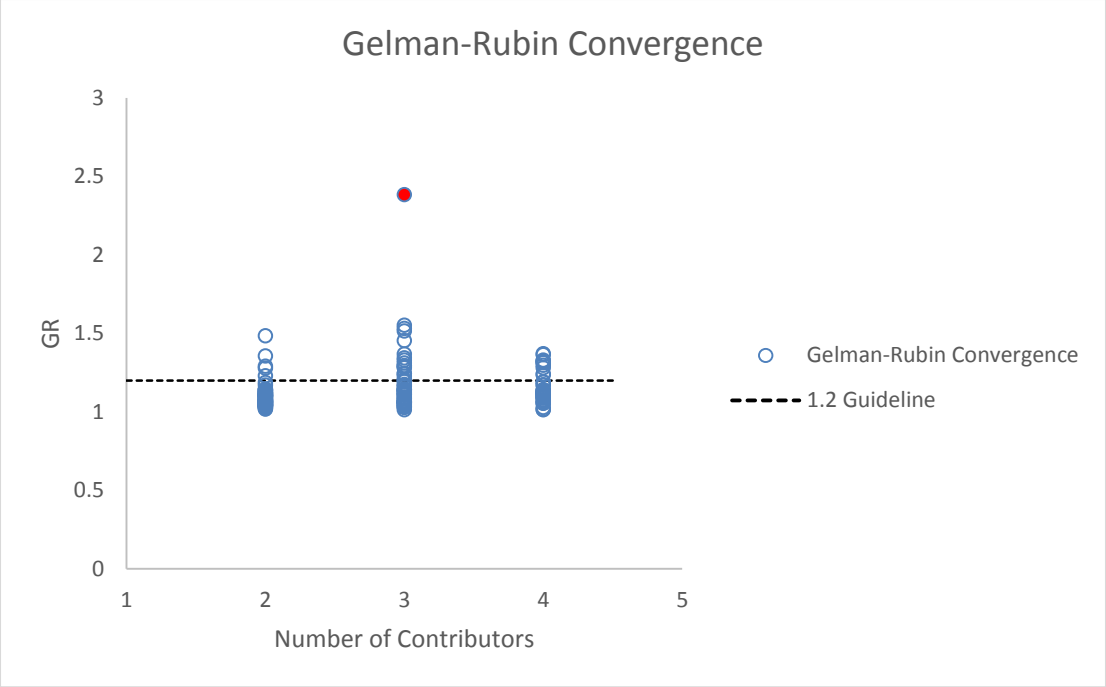


Figure D4: Plot of Gelman-Rubin convergence diagnostic (GR) compared to the true number of contributors.

Inspection of Figure D4 shows a spread of GR values for the runs, with the majority below or clustering around a value of 1.2 (the dashed line on the plot) and a maximum observed value of 2.38 obtained from one of the 3-person mixtures. The inventors of this diagnostic indicate a value below 1.2 suggests likely convergence of the MCMC chains. However, a value above this again does not necessarily mean the results are invalid. It is anticipated that for some complex mixtures this value may go above 1.2. Where a high GR value is seen, if all the remaining primary (weights, Mx and *LR*) and secondary diagnostics appear typical then there is increased confidence the results are suitable for use. If this is not the case, an option is to simply re-run the sample with either the same or an extended number of accepts.

The notable outlier at 2.38 from sample G05_M8M9F4_5-1-1_0.5_3, from the 3-person related set displayed no other obvious issues within the primary diagnostics. This sample was deconvoluted twice more, firstly using standard settings and secondly with increased iterations (x10 increase to both burn-in and MCMC). The re-run under standard settings had little impact on the GR value or the LR values for both known contributors and non-contributors. However with increased iterations the GR convergence value dropped significantly to 1.02. Under these settings all non-contributors (except F7) were excluded with LR values of 0.

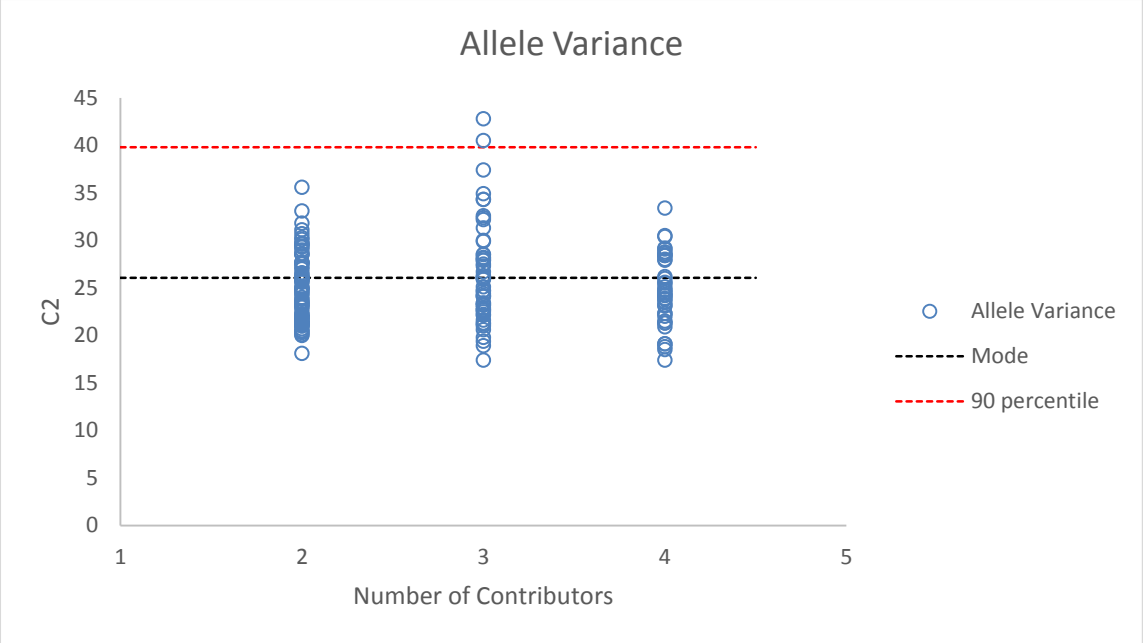


Figure D5: Plot of average allele variance proposed during the post burn-in accepts compared to the true number of contributors. The dashed lines are the mode of PBSO's prior distribution for allele variance (26.07) and the 90th percentile (39.83).

Inspection of Figure D5 shows a spread of allele variance for the samples run. The bulk of the data is around the mode of the prior distribution. The mode is a useful reference point as is a plot of the prior distribution (provided below) in order to gauge where a given posterior mean values sits. In the context of Figure D6 below, all the above data points (which range from 17.4 to 42.8) sit within the main body of this prior distribution.

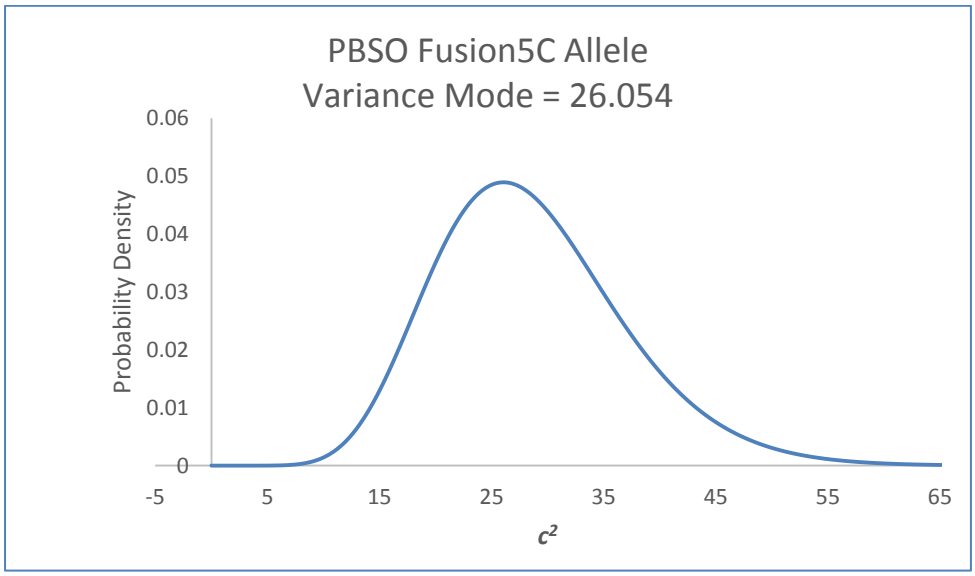


Figure D6: Plot of Allele Variance prior distribution for the Fusion® 5C kit at PBSO.

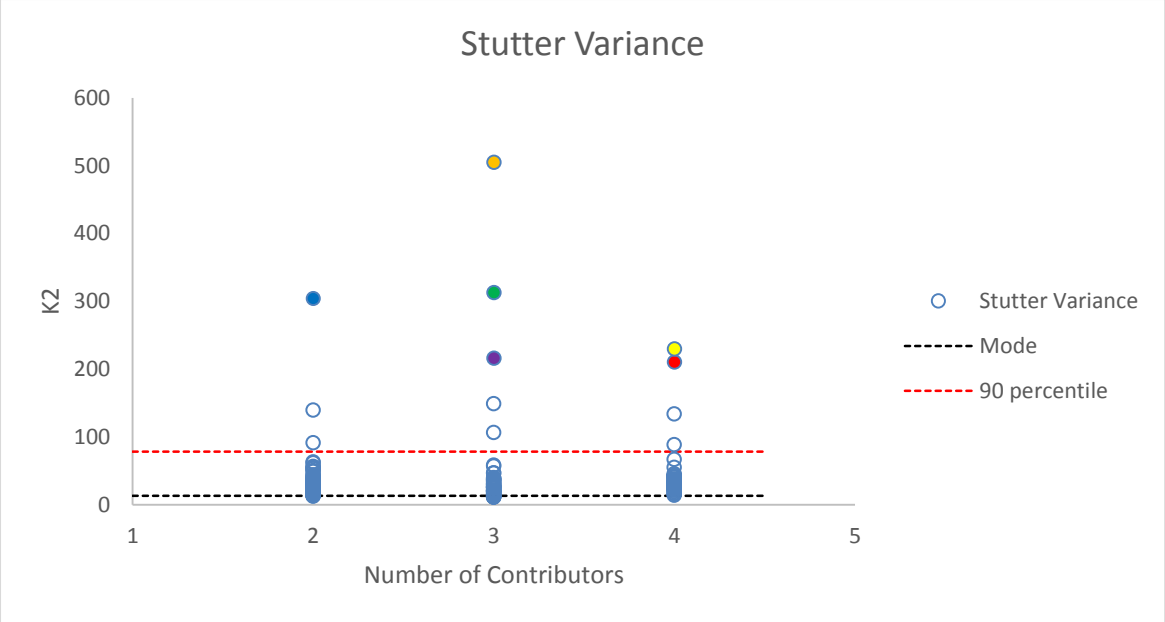


Figure D7: Plot of average stutter variance proposed during the post burn-in accepts compared to the true number of contributors. The dashed lines are the mode (13.23) and 90th percentile (78.23) of PBSO's prior distribution for stutter variance.

Inspection of Figure D7 shows a spread of stutter variance for the samples run. The majority of the samples have a value above the mode of the prior distribution. Nevertheless, all but 12 of the runs gave a value within the 90th percentile of the prior distribution, which may be a more useful point of reference with this data. Again for the PBSO a plot of the prior distribution (provided below) is more useful in order to gauge where a given posterior mean stutter variance values sits. In the context of Figure D8 below, the bulk of the data points (which range from 12.6 to 67.0) sit within the body of this prior distribution and thus could be considered within an acceptable range for this system.

Note: Please keep in mind that the variance levels in this data is high and allowing for more stochastic variation within a system has consequences during the MCMC process (see note above).

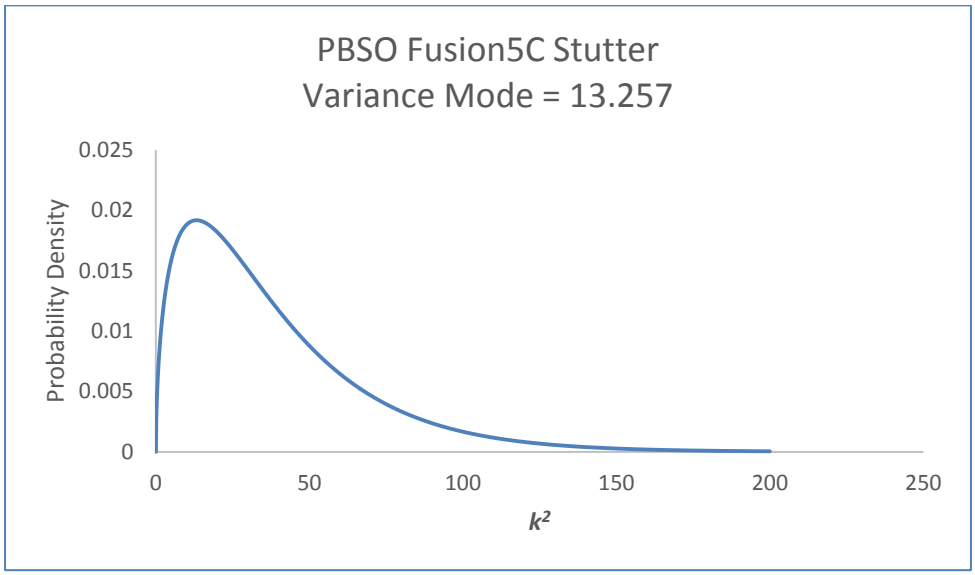


Figure D8: Plot of stutter variance prior distribution for the Fusion® 5C kit at PBSO.

Table D3 below provides some insight from the profiles as to the reasons for the highest six stutter variances observed within this data. The majority of issues were caused by peak heights exceeding saturation and lack of resolution of 1bp variant stutter. One profile had a vWA 14 allele with no apparent stutter. Variable stutter ratios have been noted at vWA 14 from different donors due to differences in sequence and LUS values.

Table D3: Stutter Variance Outliers from Figure D7.

K^2	Mixture Set	Sample	Comment
505.3	3p Unrelated	F07 5-5-1 1.0	See Ave log(Likelihood) above. Baseline noise, affecting calling (or not) of stutter notably at D16, D16 '13' > saturation 30k rfu. D12 '18' split, no stutter resolved for large 19.1 allele
313.1	3p Unrelated	E07 1-3-3 1.0	Baseline noise, affecting calling (or not) of stutter notably at D16 (and THO1). D12 '18' split, no stutter resolved for large 19.1 allele
304.2	2p Unrelated	E03 0.5ng 1-19	Possible vWA 14 issue and D16 peak (12) >saturation of 30,000 rfu
229.8	4p Unrelated	E04 4-4-1-1 1.0	*see below this sample gave an LR of 0 for F6
216.2	3p Unrelated	E06 1-1-8 1.0	Baseline noise, affecting calling (or not) of stutter notably at D18 (and D16).D18 '20' >saturation 30,000 rfu. D12 '18' split, no stutter resolved for large 19.1 allele
210.6	4p Unrelated	E04 4-4-1-1 1.0	same sample as above* but with +4p stutter labelled at vWA, D16 allele >saturation 30,000 rfu

Furthermore, inspection of the proposed mixture proportion (Mx) values from the 166 STRmix™ runs were compared to the planned Mx values during experimental set-up (data not shown). The majority of Mx values were close to the experimental design. The majority of deviations occurred as the complexity of the mixture increased (number of contributors and presence of related individuals) and at the lower template levels where the planned Mx is not exactly what is observed in the electropherograms. Mx proportions of '0' has been postulated for some minor/trace contributors in the 3 and 4 person related data sets. This is not unexpected given the masking of alleles in these profiles from major contributors.

Section E: Alternate propositions

This section covers the following standard:

> 4.1.2.1. The laboratory should evaluate more than one set of hypotheses for individual evidentiary profiles to aid in the development of policies regarding the formulation of hypotheses. For example, if there are two persons of interest, they may be evaluated as co-contributors and, alternatively, as each contributing with an unknown

individual. The hypotheses used for evaluation of casework profiles can have a significant impact on the results obtained.

In an extension of the experiment from section D of this report (as discussed for standard 4.1.2.1 above), alternate propositions were examined. One contributor is assumed (also referred to as conditioned upon) under both H_p and H_d for a sub set of the two, three and four person mixtures (see Table E1). The outcome of these alternative propositions are plotted in Figure E1 where the log(*LR*) for each individual in the database under the conditioned/assumed contributor versus the original unconditioned propositions.

The different propositions being considered are:

H_p: The DNA originated from the known individual, the database individual and *N*-2 unknown individuals

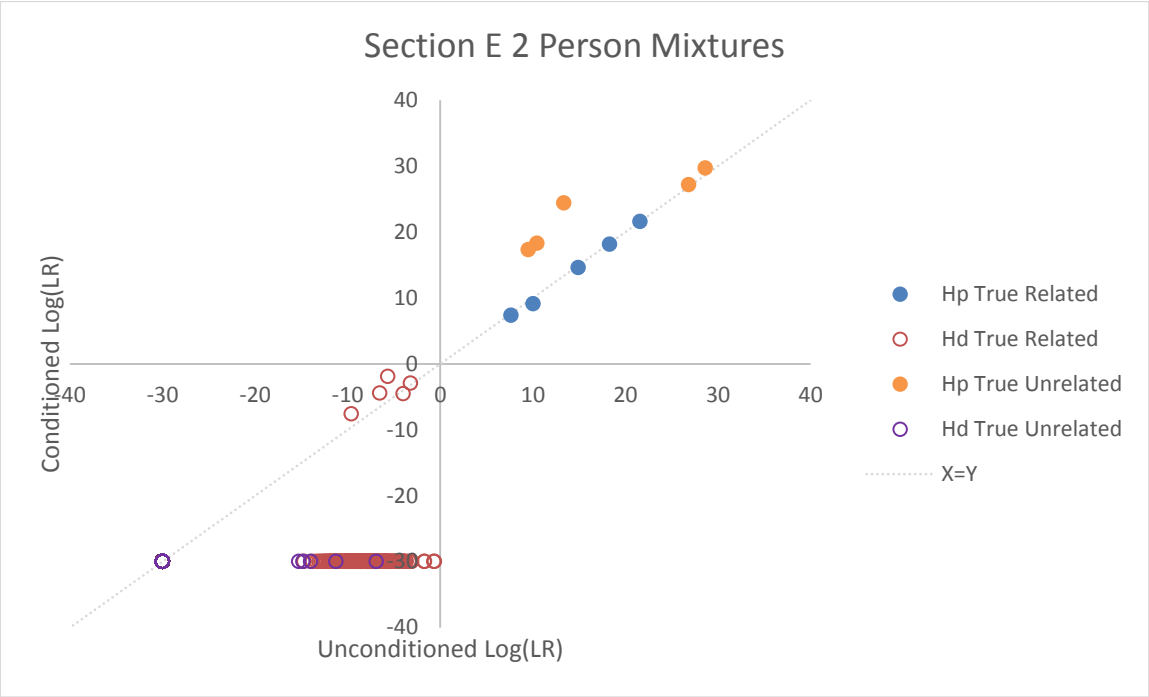
H_d: The DNA originated from the known individual and *N*-1 unknown individuals

Again the *LR*s are point estimate values using no F_{ST} and the NIST Caucasian allele frequencies.

Table E1. Mixture samples tested using alternative propositions.

Mixture Set	Major/assumed	sample
4p Unrelated	F6	C01_M10M11F5F6_1-3-3-9_0.1_2.hid.csv
		G04_M10M11F5F6_1-1-3-6_0.1_1.hid.csv
		A02_M10M11F5F6_1-1-3-6_0.5_3.hid.csv
		D06_M10M11F5F6_1-1-3-6_0.25_2.hid.csv
		G04_M10M11F5F6_1-1-3-6_1.0_1.hid.csv
4p Related	F7	C02_M1M2M3F7_1-1-3-6_0.1_1.hid.csv
		D04_M1M2M3F7_1-1-3-6_0.25_1.hid.csv
		G08_M1M2M3F7_1-1-3-6_1.0_3.hid.csv
		D06_M1M2M3F7_1-1-3-6_0.5_1.hid.csv
		F02_M1M2M3F7_1-3-3-9_0.1_1.hid.csv
3p Unrelated	M6	E04_M6M7F3_6-3-1_0.1_2.hid.csv
		D01_M6M7F3_6-3-1_0.25_1.hid.csv
		A03_M6M7F3_5-5-1_0.25_2.hid.csv
		E05_M6M7F3_5-5-1_0.5_3.hid.csv
		B06_M6M7F3_5-5-1_0.1_3.hid.csv
3p Related	M8	E05_M8M9F4_5-5-1_0.1_1.hid.csv
		A04_M8M9F4_6-3-1_0.1_1.hid.csv
		E02_M8M9F4_5-5-1_0.25_1.hid.csv
		A01_M8M9F4_6-3-1_0.25_1.hid.csv
		A04_M8M9F4_6-3-1_0.5_1.hid.csv
2p Unrelated	F2	A03_F2M5_5-1_0.1_2.hid.csv
		F02_F2M5_2.5-1_0.1_2.hid.csv
		A06_F2M5_5-1_0.25_2.hid.csv
		E05_F2M5_2.5-1_0.25_1.hid.csv

		E02_F2M5_2.5-1_0.5_1.hid.csv



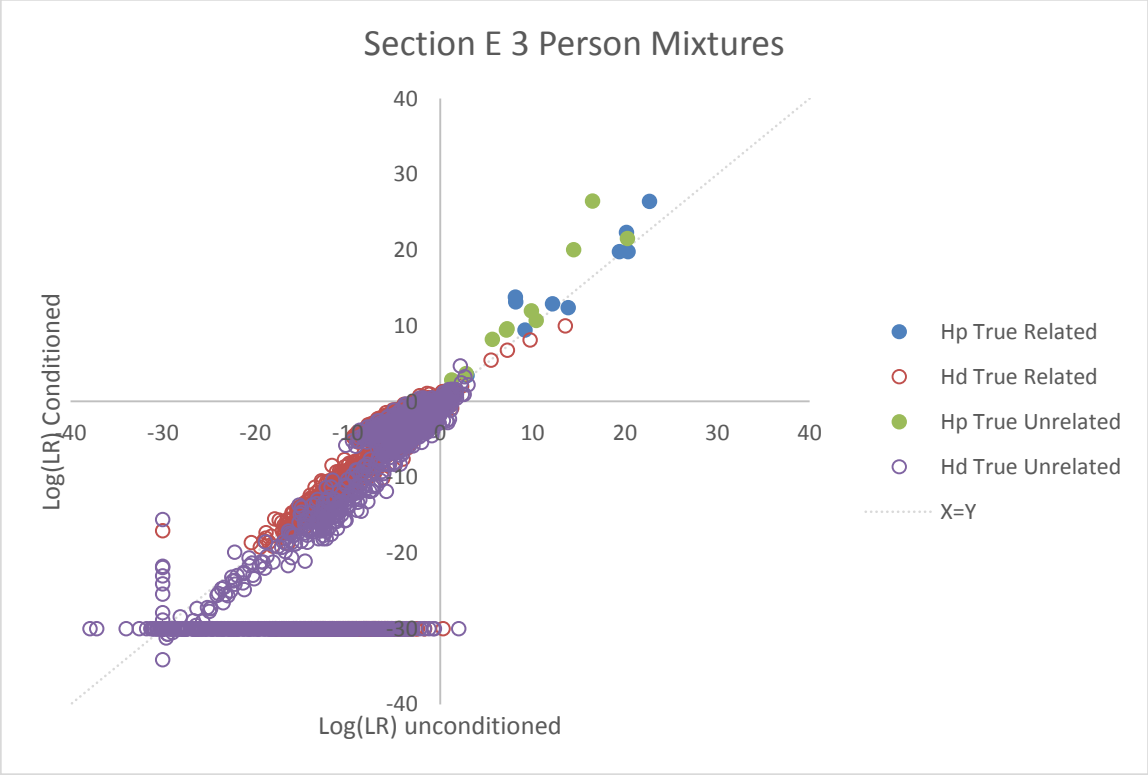


Figure E1 : Comparison of Log(*LR*) obtained from deconvolutions performed with and without assuming/conditioning profiles.

Positive values for Hd true were obtained from the "related" data series due to F7 matches.

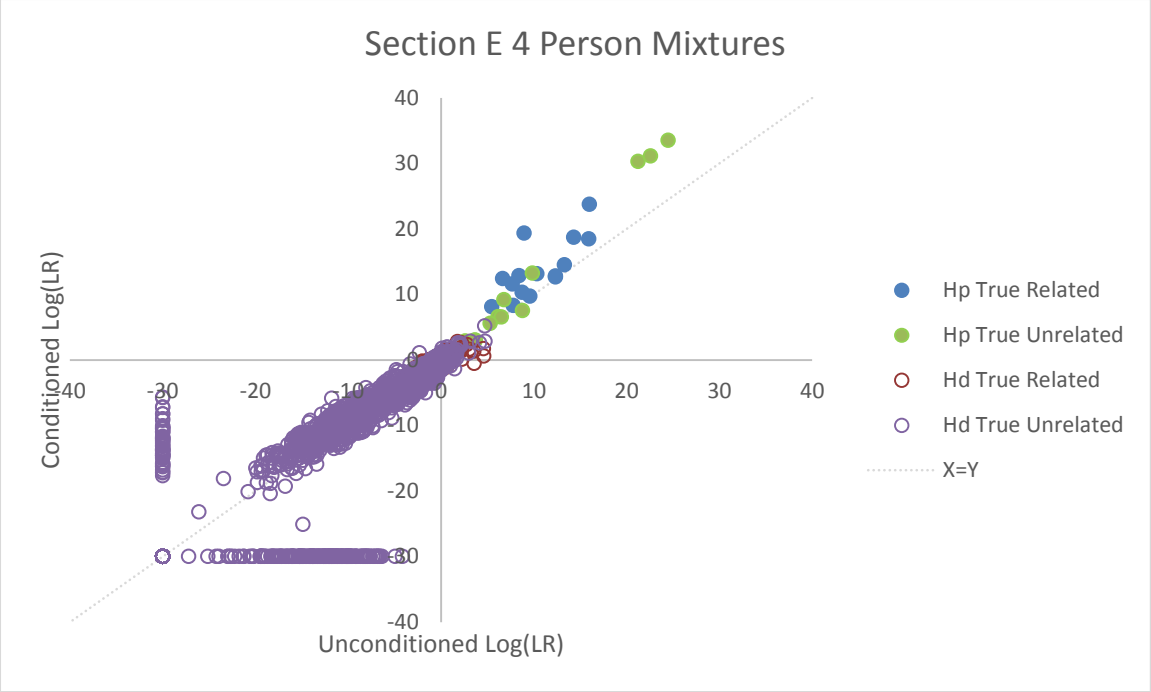


Figure E1 (continued): Comparison of Log(*LR*) obtained from deconvolutions performed with and without assuming/conditioning profiles.

Inspection of the plots in Figure E1 indicates that, in many instances there is an increase in the *LR* of the known contributors when a conditioning profile is used (known contributors under both H_p and H_d) compared to when no conditioning information is available. This can be seen by the blue and green circles being above the dashed (X=Y) line. It is also apparent in these plots that the use of a conditioning profile may decrease the *LR* or lead to outright exclusion (*LR* of 0 inserted at Log(*LR*) =-30) of non-contributors. This can be seen by many of the purple and red circles being below the dashed line. However in a few samples, notably D06 1-1-3-6 at 0.25ng in the 4 person unrelated data set, conditioning on one contributor has given some low log(LR) values (maximum 10^{-7}) for known non contributors that were previously excluded. In summary, where it is appropriate to use conditioning profiles, the addition of correct conditioning profile information improves the performance of STRmix™ in its ability to differentiate between known contributors and non-contributors.

Section F: Assigning number of contributors

This section covers the following standard:

> 4.1.6.4. If the number of contributors is input by the analyst, both correct and incorrect values (i.e., over- and under-estimating) should be tested.

The effect of the uncertainty in the number of contributors within STRmix™ has previously been reported for a number of profiles with *N* and *N*+1 assumed contributors, where *N* is the number of contributors [4, 5]. The inclusion of an additional contributor beyond that present in the profile had the effect of lowering the *LR* for trace contributors within the profile. STRmix™ adds the additional

(unseen) profile at trace levels which interacts with the known trace contribution, diffusing the genotype weights and lowering the *LR*. There was no significant effect on the *LR* of the major or minor contributor within the profiles.

The effect was tested by both increasing and decreasing the number of contributors compared with the known (*N*+1 and *N*-1 trials). The true number of contributors to a profile is always unknown. Analysts are likely to add contributors in the presence of an artifact, high stutter, or forward stutter peaks. The assumption of one fewer contributor than that actually present may be made when contributors are at very low levels and dropping out (or visible below the analytical threshold), in constructed profiles where DNA is from individuals with similar profiles at the same concentrations, or family scenarios, such as DNA from a father, mother and their child where the child was the minor contributor.

Addition of one contributor

A selection of two and three person mixtures were interpreted as three and four person profiles, respectively. (See Table F1)

Table F1. Selection of two and three person mixtures.

Mixture Set	Run in STRmix as	Sample
2p Unrelated	3 (N+1)	E04 F2M5 1-5 0.25-2
		F01 F2M5 1-5 0.5-3
		G02 F2M5 2.5-1 0.1-3
2p Related	3 (N+1)	G03 F1M4 10-1 0.5-3
		F06 F1M4 10-1 0.25-2
3p Unrelated	4 (N+1)	C05 M6M7F3 1-1-1 0.1-2
		C04 M6M7F3 1-1-8 0.5-2
		D01 M6M7F3 6-3-1 0.25-1
		E05 M6M7F3 5-5-1 0.5-3
		A02 M6M7F3 1-1-8 0.25-3

The *LR* for both the known contributors and 500+ known non-contributors (as for the specificity and sensitivity studies, Section D) were calculated. The LRs calculated for the known contributors and known non-contributors under the assumption of *N* and *N*+1 contributors were compared in plots in Figure F1.

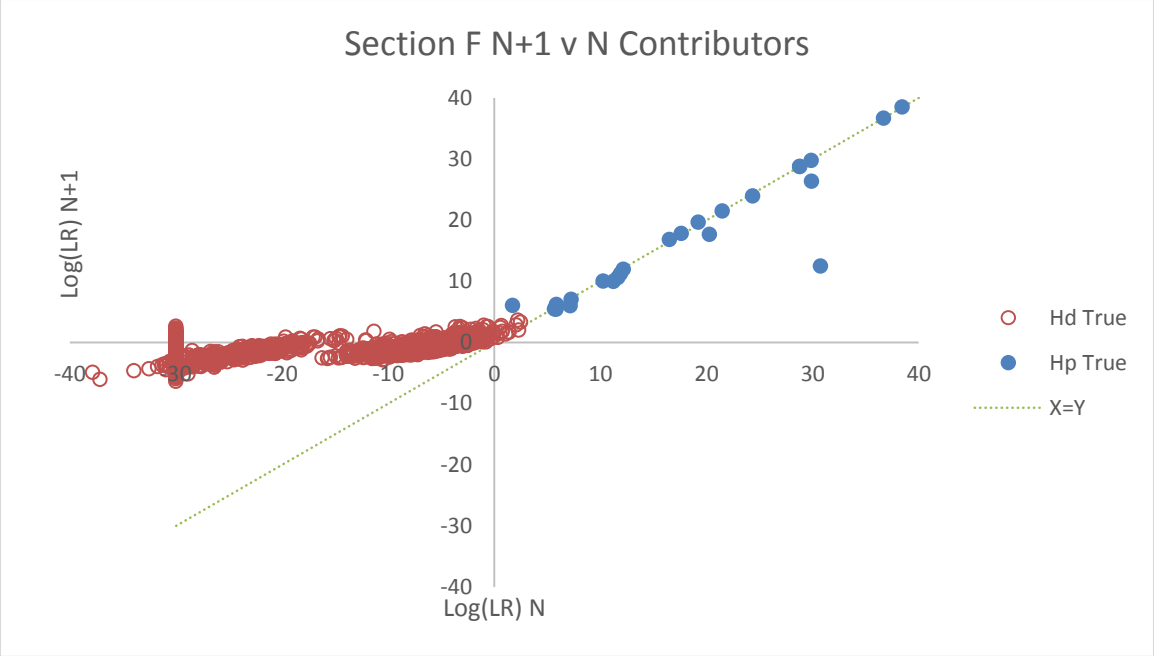


Figure F1: Effect of over-estimating the number of contributors.

The above plot illustrates that over estimating the number of contributors can reduce the LR of true contributors, notably for minor, low level contributors. Increasing the number of contributors can increase the number of positive LRs for known non contributors however, none of the non-contributors gave exclusions (LRs of 0) plotted as Log(LR)= -30.

Subtraction of one contributor

A selection of three and four person mixtures were deconvoluted with the number of contributors set to two and three respectively (see Table F2)

Table F2. Selection of three and four person mixtures.

Mixture Set	Run in STRmix as	Sample
3p Related	2 (N-1)	A05 M8M9F4 1-1-1 0.1-3
		E05 M8M9F4 5-5-1 0.1-1
		B05 M8M9F4 1-3-3 0.1-1
		A04 M8M9F4 6-3-1 0.1-1
4p Related	3 (N-1)	A02 M1M2M3F7 4-4-1-1 0.1-2
		B06 M1M2M3F7 4-4-1-1 0.5-2
		F02 M1M2M3F7 1-3-3-9 0.1-1
		A04 M1M2M3F7 4-4-1-1 0.25-1
4p Unrelated	3 (N-1)	D04 M10M11F5F6 4-4-1-1 0.1-1
		G04 M10M11F5F6 1-1-3-6 0.1-1

The *LR* for both the known contributors and 500+ known non-contributors (as for the specificity and sensitivity studies, Section D) were calculated. The LRs calculated the known contributors and

known non-contributors under the assumption of *N* and *N*-1 contributors were compared in plots in Figure F2.

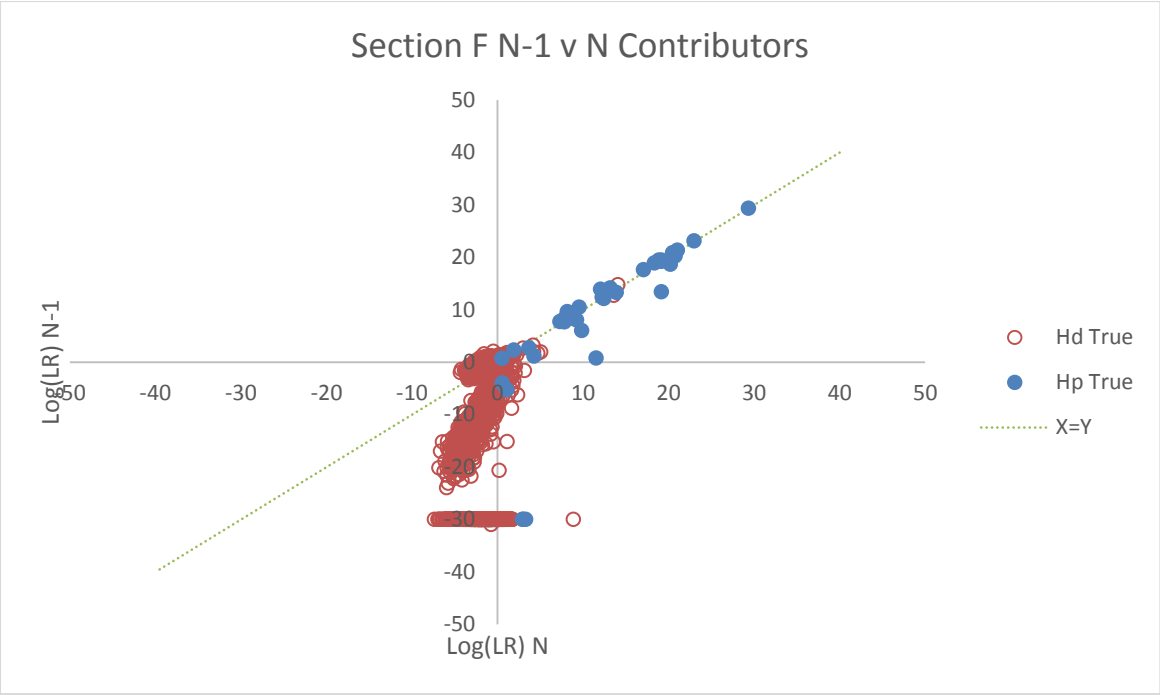


Figure F2: Effect of under-estimating the number of contributors.

Inspection of Figure F2 shows little movement of the LR for some known contributors (the major/intermediate components) but a false exclusion for some (the minor/trace components) when run as *N*-1. This is not unexpected as STRmix™ has to spread the genotype combinations across fewer proposed contributors and hence some genotype combinations of the true contributors aren't considered. The log(LR) values for the known contributors to two 4 person and two 3 person mixtures are shown in Table F3. Those which become exclusions are highlighted in yellow. It should be noted, that the Hd true LR for individual F7 in 3p related mixtures was typically >1, indicating that F7 is likely related to the true contributors.

Table F3: Log(*LR*) values for some three and four person mixtures assuming *N* and *N*-1 contributors.

Set	Sample	Contributor	Mx	Log(LR) N	Log(LR) N-1
4p U	D04	M10	4	22.99605	23.10487
		M11	4	7.29127	7.735178
		F5	1	4.284452	1.130836
		F6	1	1.080431	-5.31092
4p U	G04	M10	1	3.656532	2.699024
		M11	1	0.550943	-3.9919
		F5	3	9.867841	6.000755
		F6	6	12.25725	12.35735
3p R	B05	M8	1	3.297621	-30

		M9	3	19.17925	19.46949
		F4	3	20.77397	20.207
3p R	E05	M8	5	21.04901	21.34725
		M9	5	20.20325	18.66974
		F4	1	2.942358	-30

Section G: Drop-in

This section covers the following standard:

4.1.8 Allele drop-in

Drop-in has been observed with the use of PowerPlex® Fusion with 30 cycle amplification. Drop-in parameters have been included in the Palm Beach Fusion DNA profiling kit in STRmix™ and were defined during Model Maker. Drop-in peaks either greater or less than the 300 RFU drop-in cap were added to single source profiles and interpreted by STRmix™. Table G1 summarizes the total LR calculated by STRmix™ for the single source samples with and without the drop-in alleles.

Table G1. Total LR values with and without drop-in alleles.

	Below Dropin Threshold (<300RFU)							
Sample Name	Drop in Peak RFU	LR without dropin	LR (Caucasian)	LR without dropin	LR (African American)	LR without dropin	LR (Hispanic)	
16-000574_3_4	135	2.30E+29	2.30E+29	1.31.E29	1.31.E29	4.38E+31	4.38E+31	* Added at Heterozygous marker
16-575 Item 4_2	147	4.58E+26	4.58E 26	2.65E+30	2.65E+30	1.70E+27	1.70E+27	* Added at Homozygous marker
15-000071_4_1	120	1.58E+28	1.58E+28	8.70E+28	8.70E+28	8.52E+27	8.52E+27	* Added at Homozygous marker

	Above Dropin Threshold (>300RFU)							
Sample Name	Drop In Peak RFU	LR without dropin	LR (Caucasian)	LR without dropin	LR (African American)	LR without dropin	LR (Hispanic)	
14-000072_3_3	315	1.75E+27	Software did not interpret sample	1.62E+30	Software did not interpret sample	3.83E+27	Software did not interpret sample	* Added at Heterozygous marker
13-000074_3_4	320	4.62E+27	Software did not interpret sample	1.40E+31	Software did not interpret sample	4.94E+28	Software did not interpret sample	* Added at Heterozygous marker
13-000071_4_6	330	3.18E+28	0	1.05E+31	0	3.80E+29	0	* Added at Homozygous marker

Table G1 illustrates that total LR did not change when the drop-in peak is within model parameters. For alleles that are outside of model parameters, a LR of zero was returned or the interpretation in STRmix™ did not progress as the profile can no longer be explained by one contributor.

Section H: Forward and reverse stutter

This section covers the following standard:

4.1.9 Forward and reverse stutter

STRmix™ implements a "per allele" back stutter model. This is alternatively based on the longest uninterrupted stretch (LUS) of common repeats in the allele or the allele designation itself. STRmix™ can also implement either a per allele or per locus forward stutter model. Stutter peak labels (-1 and +1 stutters) are retained at analysis and within the STRmix™ input file. The modelling of stutter peaks may be seen in the interpretation of single source profiles where stutter peaks are retained at interpretation. As part of the Markov Chain Monte Carlo (MCMC) process they are considered as

alleles in the genotype combination proposed for a given iteration but those combinations result in very low probabilities, and are not accepted, therefore receiving no weight. In mixed DNA profiles, where the minor contributor is of similar height as the stutter peaks, the stutter peaks start to be considered as minor alleles. This is as expected.

Forward stutter parameters have been included in the Palm Beach Fusion DNA profiling kit in STRmix™ and were defined during Model Maker. Forward stutter alleles either greater or less than the 15% forward stutter cap were added to single source profiles and interpreted by STRmix™. Table H1 summarizes the total LR calculated by STRmix™ for the single source samples with and without the forward stutter allele added.

Table H1. Total LR values with and without forward stutter allele.

Addition of Forward stutter (N+4) less than 15%

		Below 15% N+4						
Sample Name	N+4 RFU	LR without stutter peak	LR (Caucasian)	LR without stutter peak	LR (African American)	LR without stutter peak	LR (Hispanic)	
13-000071-4-6	2000 (13%)	3.18E+28	3.18E+28	1.05E+31	1.05E+31	3.80E+29	3.80E+29	* Added at Homozygous marker
13-000074_3_4	2700 (13%)	4.62E+27	4.62E+27	1.40E+31	1.40E+31	4.94E+28	4.94E+28	* Added at Heterozygous marker

Addition of Forward stutter (N+4) greater than 15%

		Greater than 15% N+4						
Sample Name	N+4 RFU	LR without stutter peak	LR (Caucasian)	LR without stutter peak	LR (African American)	LR without stutter peak	LR (Hispanic)	
16-000574_3_4	2700 (17%)	4.62E+27	0	1.40E+31	0	4.94E+28	0	* Added at Homozygous marker
16-575 Item 4 _2	753 (15.1%)	4.58E+26	Software did not interpret sample	2.65E+30	Software did not interpret sample	1.70E+27	Software did not interpret sample	* Added at Heterozygous marker
14-000072_3_3	3200 (17%)	1.75E+27	0	1.62E+30	0	3.83E+27	0	* Added at Homozygous marker

Table H1 illustrates that the total LR did not change when the forward stutter peak is within model parameters. For forward stutter alleles that are outside of model parameters, a LR of zero was returned or the interpretation in STRmix™ did not progress as the profile can no longer be explained by one contributor.

Section I: Intra locus peak height

This section covers the following standard:

4.1.10. Intra-locus peak height variance

STRmix™ models the variability of single peaks. The variance of this model is determined by directly modelling laboratory data. This is undertaken within STRmix™ using the Model Maker function. Traditionally heterozygous balance (H*b*) for a STR typing kit is investigated. Heterozygous balance can be thought of as the variability of two alleles at a heterozygous locus. A plot of log (H*b*) versus average peak height (APH) of a locus demonstrates that the variability in H*b* decreases as APH increases. The performance of Model Maker is checked by plotting the bounds informed by the Model Maker Results (refer to Estimation of STRmix™ Parameters for Palm Beach County Sheriff's Office report for further details).

The plot of log(H*b*) versus APH and the expected 95% bounds (plotted as dotted lines) calculated by $\pm \sqrt{2}\,\mathrm{x}1.96\mathrm{x}\,\frac{\sqrt{c^2}}{APH}$ where $c^2 = 27.74$, the 50$^{\text{th}}$ percentile from the gamma distribution, determined for the Palm Beach PowerPlex® Fusion data during Model Maker. The plot of log(H*b*) versus APH is given in Figure I1.

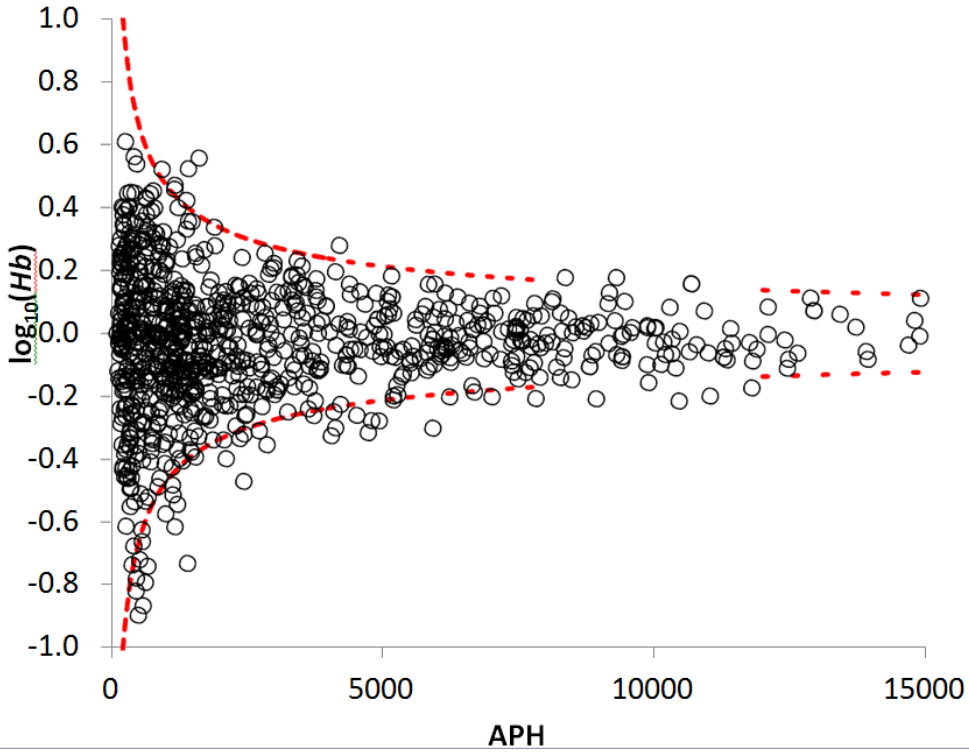


Figure I1. Log(HB) versus SPH for the combined dataset of single source profiles for Palm Beach PowerPlex® Fusion data.

Section J: Inter-Locus peak heights

This section covers the following standard:

4.1.11 Inter-locus peak height variance

Inter locus peak variance is modelled in STRmix™ using locus specific amplification efficiencies (LSAE). The LSAE model reflects the observation that even after template DNA amount, degradation and variation in peak height within loci are modelled, the peak heights between loci are still more variable than predicted. The variance of this model is determined by directly modelling laboratory data. LSAE values for each STRmix™ interpretation appear within the results. We can demonstrate the relationship of LSAE values to average peak heights (APH) in a graphical format. The LSAE values should mimic the average peak heights of the locus. This is demonstrated in Figures J1-J10.

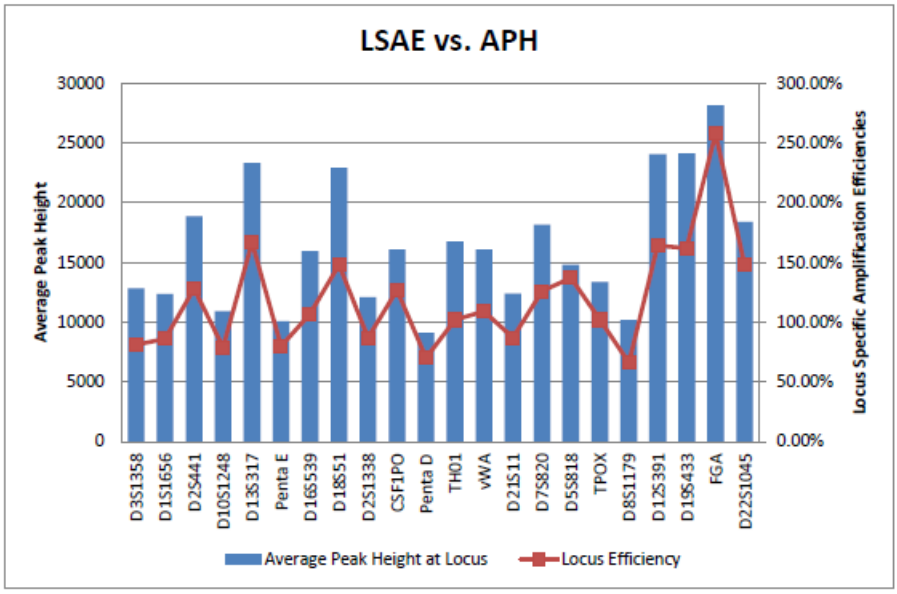


Figure J1. Plot of APH and LSAE value for each locus for sample one with APH approaching the saturation threshold of 30000 rfu.

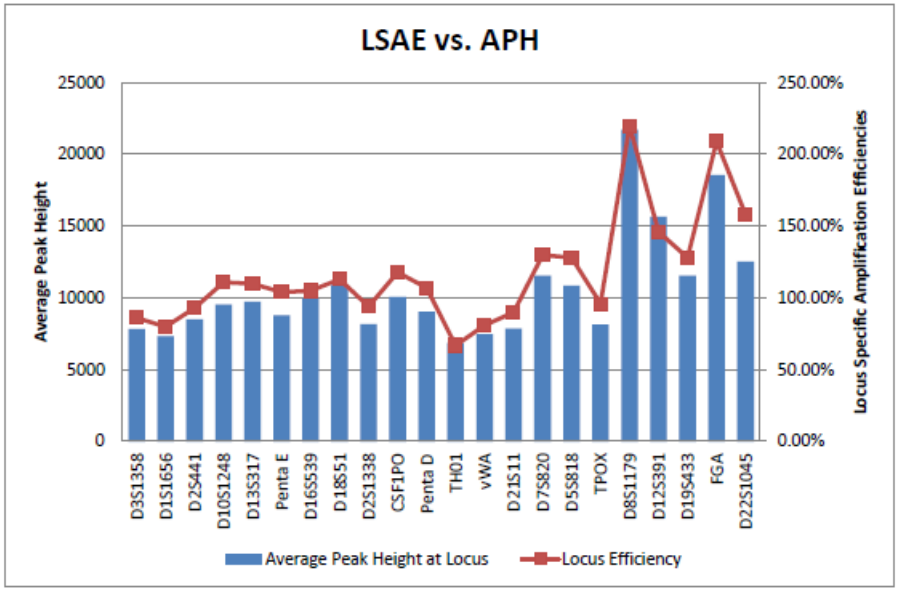


Figure J2. Plot of APH and LSAE value for each locus for sample two.

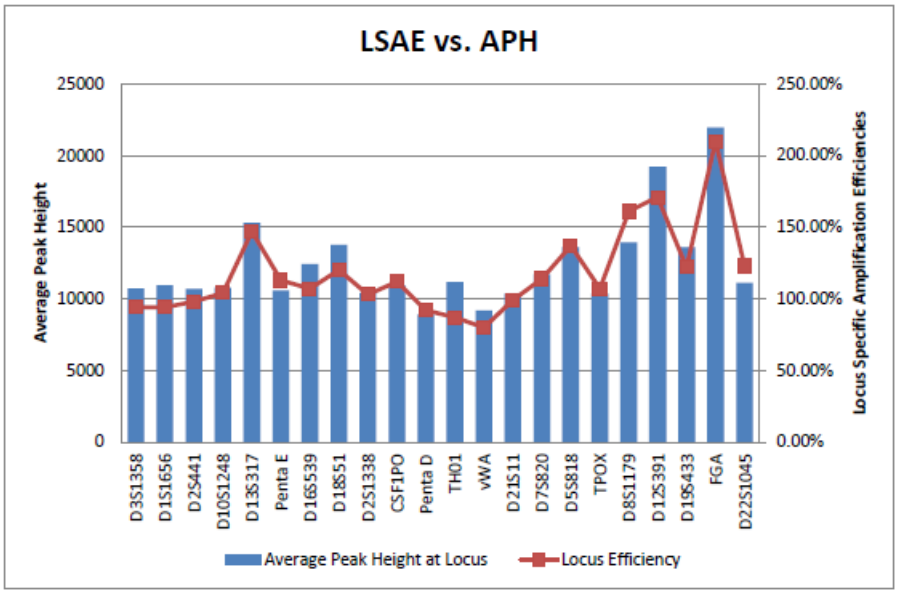


Figure J3. Plot of APH and LSAE value for each locus for sample three.

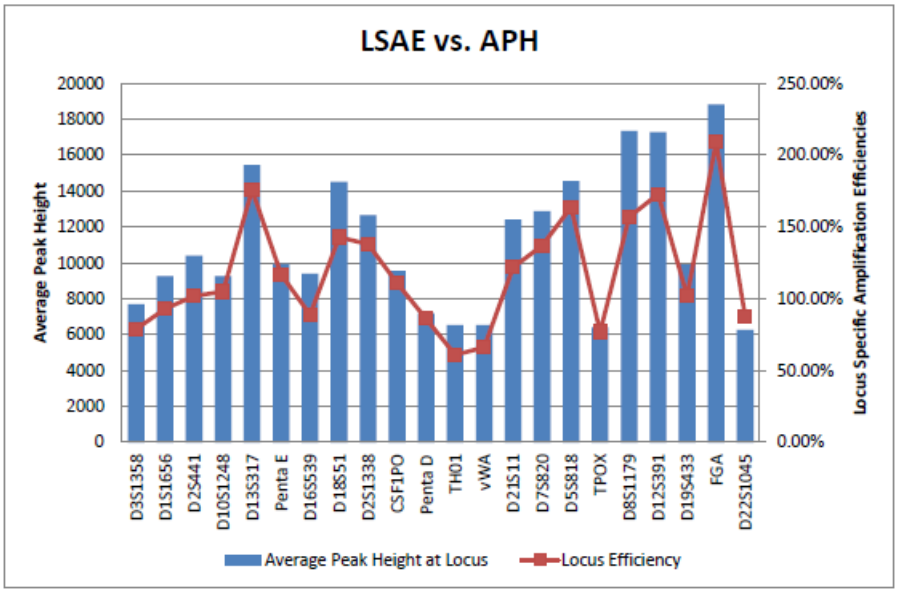


Figure J4. Plot of APH and LSAE value for each locus for sample four.

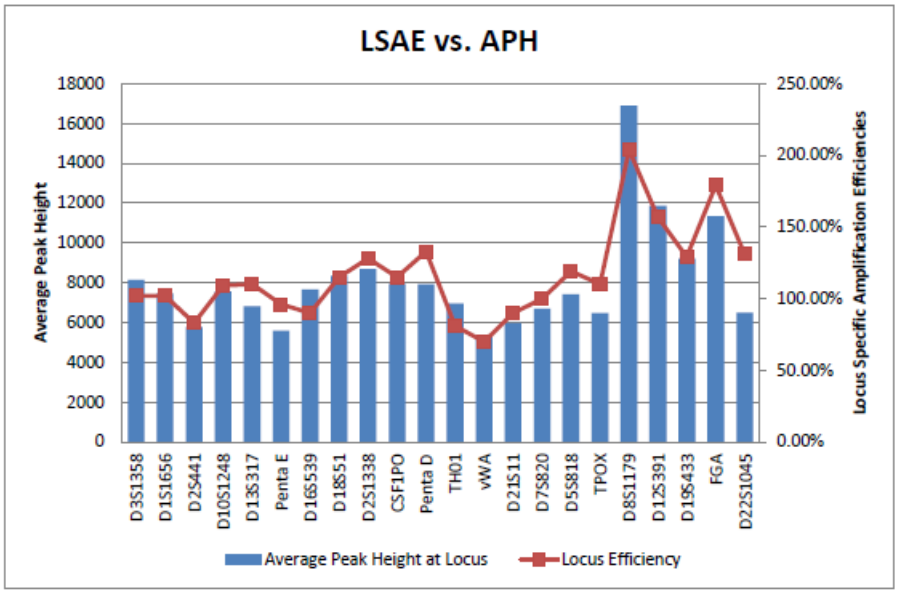


Figure J5. Plot of APH and LSAE value for each locus for sample five.

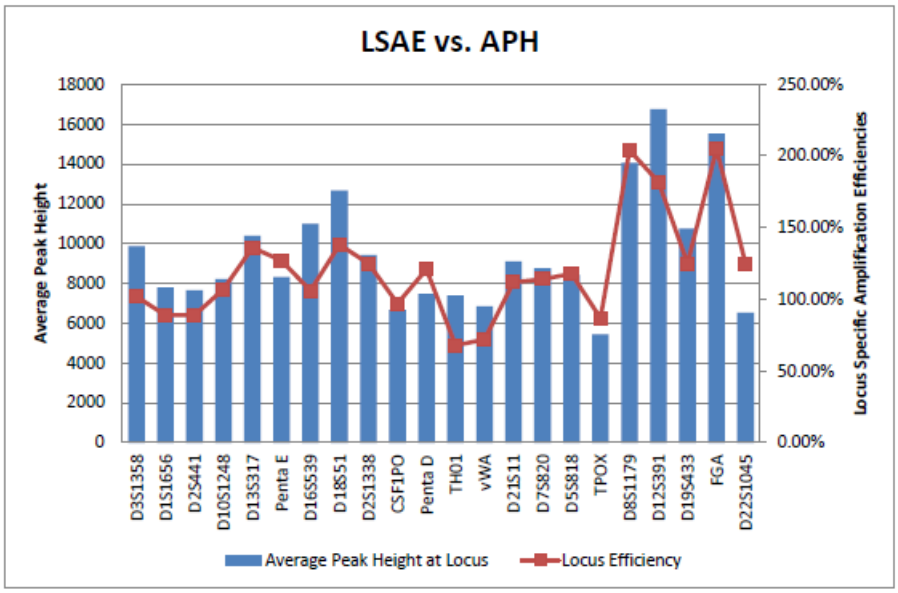


Figure J6. Plot of APH and LSAE value for each locus for sample six.

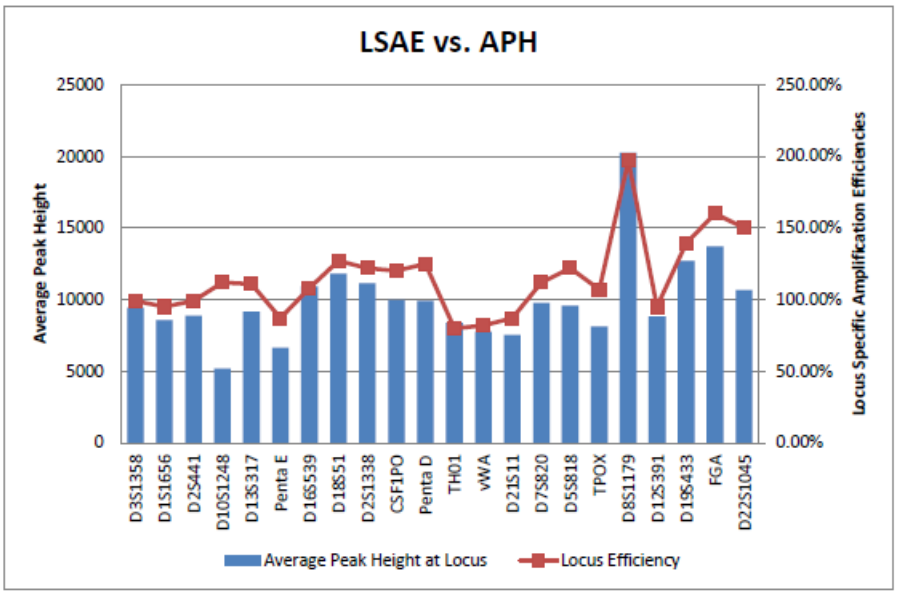


Figure J7. Plot of APH and LSAE value for each locus for sample seven.

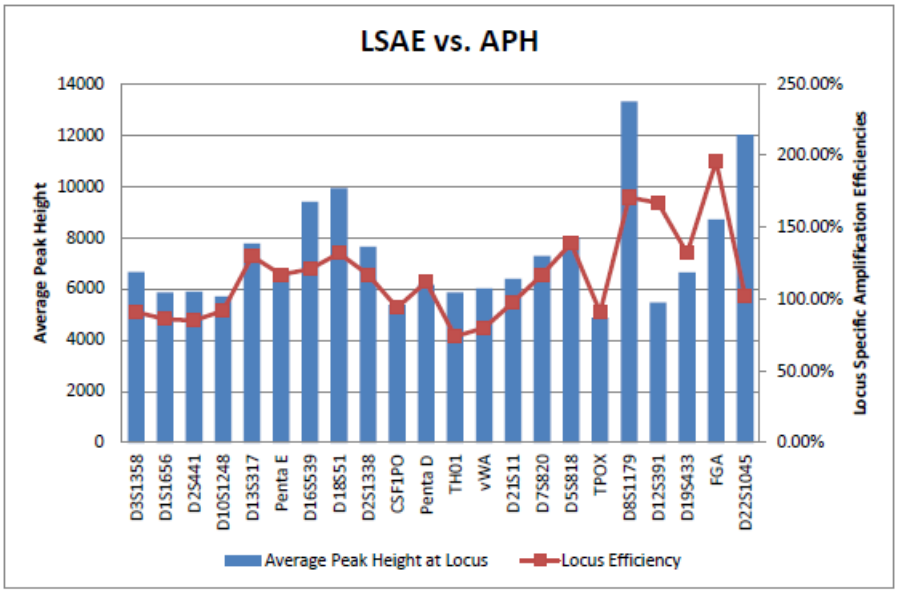


Figure J8. Plot of APH and LSAE value for each locus for sample eight.

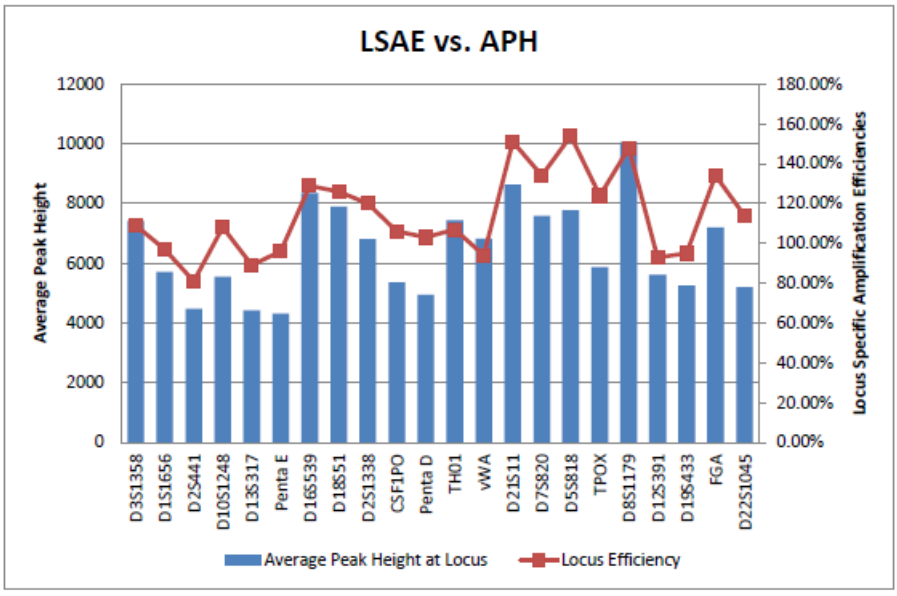


Figure J9. Plot of APH and LSAE value for each locus for sample nine.

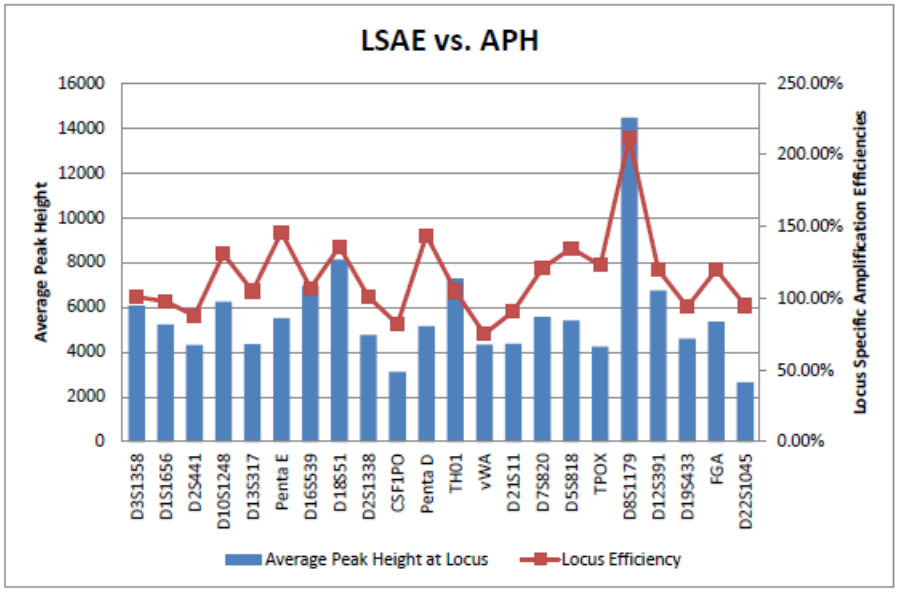


Figure J10. Plot of APH and LSAE value for each locus for sample ten.

Section K: Challenge testing

This section covers the following standard:

> 4.1.14 Additional challenge testing (e.g. the inclusion of non-allelic peaks such as off ladder (OL) peaks that may results from bleed through or spikes in the typing results).

STRmix™ requires that only numeric values are retained within the input file. Any values that are not numeric (such as OL alleles not removed at analysis) will cause STRmix™ to halt the interpretation. The presence of a non-allelic peak (or peaks) that has sized within an allelic bin position and is retained within the input file can cause a number of results depending on the scenario. These include:

- An exclusionary LR. If the artifact is modelled as having originated from the person of interest (for example if the peak is of a similar height to the alleles corresponding to the person of interest in a mixed DNA profile) this may result in exclusion.
- No effect. If drop-in is observed within a laboratory, the artifact may be modelled as a drop-in peak if it is less than the drop-in height threshold.
- Failure to interpret. If an artifact within an allelic bin is retained in a profile it may artificially increase the minimum number of contributors with in the profile. For example, an artifact at a heterozygous locus in a single source profile (not modelled as stutter or drop-in) will increase the minimum number of contributors by one. STRmix™ will not precede assuming only one contributor.

Each of these expected outcomes was demonstrated by editing an input file and calculating a LR within STRmix™. A summary of the effect is listed in Table K1.

Table K1. Summary of effects obtained from various scenarios where input files included one or more peaks which should have been removed.

Addition of artifact peak (non-allelic / out of bin)

Sample Name	Artifact Peak added	LR without OL allele	LR (Caucasian)	LR without OL allele	LR (African American)	LR without OL allele	LR (Hispanic)	
13-000073_Item3_19	TH01 OL @ 105 bp	1.69E+28	1.18E+16	1.49E+29	2.08E+16	2.85E+27	4.41E+15	* Added at Heterozygous marker, stat stopped at TH01
15-000071_4_1	D3 OL @ 140 bp	1.58E+28	0	8.70E+28	0	8.52E+27	0	* Added at Heterozygous marker; stat stopped at D3
14-000072_3_3	D16 OL @ 108.95 bp	1.75E+27	1.38E+06	1.62E+30	3.11E+06	3.83E+27	2.14E+06	* Added at Homozygous marker; stat stopped at D16
13-000074_3_4	TH01 OL (spike) @ 92.79 bp	4.62E+27	7.88E+14	1.40E+31	1.59E+16	4.94E+28	1.74E+15	* Added at Homozygous marker; stat stopped at TH01
13-000071_4_6	TPOX OL @ 424 bp	3.18E+28	4.55E+21	1.05E+31	4.38E+23	3.80E+29	2.96E+22	* Added at Heterozygous marker; stat stopped at TPOX

Addition of artifact peak ("allele" / within bin)

Sample Name	Artifact Peak added	LR without additional allele	LR (Caucasian)	LR without additional allele	LR (African American)	LR without additional allele	LR (Hispanic)	
13-000073_Item3_19	D19 "13.2" = 118 RFU	1.69E+28	1.69E+28	1.49E+29	1.49E+29	2.85E+27	2.85E+27	* Added at Heterozygous marker; Hp @ D19 <<< 1
15-000071_4_1	D3 "10" = 494 RFU	1.58E+28	Software did not interpret sample	8.70E+28	Software did not interpret sample	8.52E+27	Software did not interpret sample	* Added at Heterozygous marker
14-000072_3_3	D8 "15.1" = 134 RFU	1.75E+27	1.75E+27	1.62E+30	1.62E+30	3.83E+27	3.83E+27	* Added at Heterozygous marker; Hp @ D8 <<< 1
13-000074_3_4	D19 "9.2" = 322 RFU	4.62E+27	0	1.40E+31	0	4.94E+28	0	* Added at Homozygous marker; Hp @ D19 = 0
13-000071_4_6	D3 "9" = 280 RFU	3.18E+28	3.18E+28	1.05E+31	1.05E+31	3.80E+29	3.80E+29	* Added at Homozygous marker; Hp @ D3 <<< 1

Section L: Casework profiles

This section covers the following standards:

4.2 Laboratories with existing interpretation procedures should compare the results of probabilistic genotyping and of manual interpretation of the same data, notwithstanding the fact that probabilistic genotyping is inherently different from and is not directly comparable to binary interpretation. The weights of evidence that are generated by these two approaches are based on different assumptions, thresholds, and formulae. However, such a comparison should be conducted and evaluated.

4.2.1 The laboratory should determine whether the results produced by the probabilistic genotyping software are intuitive and consistent with expectations based on non-probabilistic mixture analysis methods.

4.2.1.1 Generally, known specimens that are included based on non-probabilistic analyses would be expected to also be included based on probabilistic genotyping.

4.1.7 Partial profile, to include the following:

4.1.7.2 DNA degradation

4.1.7.3 Inhibition

Eleven non-probative (NP) samples comprising of degraded, inhibited, and low level samples and covering the range of scenarios where the person of interest (POI) was considered to be excluded, inconclusive, and included were re-interpreted by STRmix™ and calculated LRs were plotted in Figure L1.

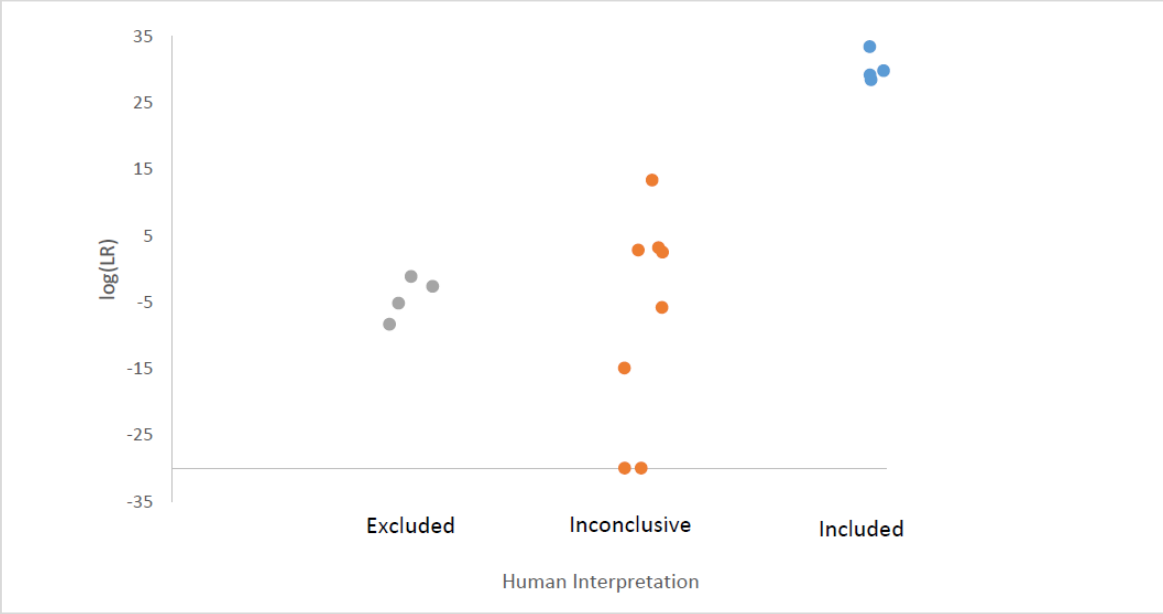


Figure L1. Plot of the Log(LR) against the manual (human) interpretation for non-probative samples.

Section M: Precision

This section covers the following standard:

4.1.13. Sensitivity, specificity and **precision,** as described for developmental validation

Refer to section D above for details of sensitivity and specificity tests.

Precision

The MCMC process is used to generate the weights within STRmix™ for different genotype combinations. This is a sampling procedure and therefore the weights will vary slightly between each run. The variability in LRs between replicate interpretations has previously been explored [9]. The MCM process was shown to be a small source of variability compared with other lab variables. The variability due to the size of the allele frequency database and the MCMC process is taken into account within STRmix™ 2.4 using the highest posterior density (HPD) method [10-12] (a type of confidence interval).

The extent of the STRmix™ run variability was investigated by interpreting three DNA profiles (from Section D of the validation) where there was ambiguity in the genotype combinations. Each DNA mixture was interpreted by STRmix™ five times using the following number of MCMC iterations 50,000, 500,000, and 5,000,000. Plots of the log(LR) generated from STRmix™ versus the MCMC iteration were generated using either the Palm Beach Fusion Caucasian, African American, or Hispanic population data (see Figures M1 –M3).

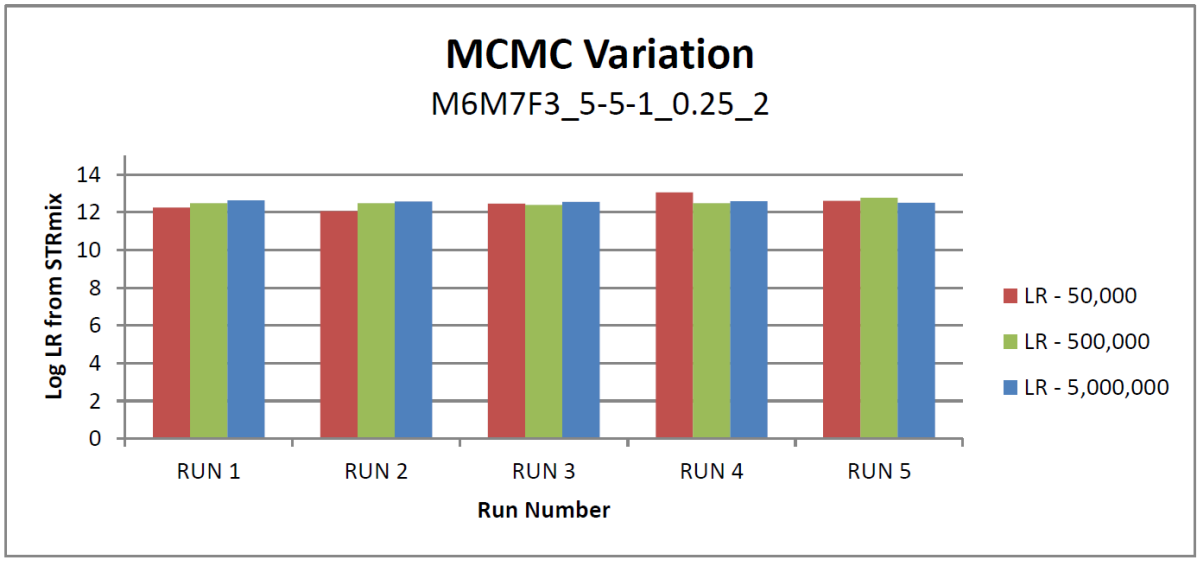


Figure M1. Plot of log(LR) versus MCMC iterations for sample M6M7F3_5-5-1_0.25_2 in the Palm Beach Fusion Caucasian population data

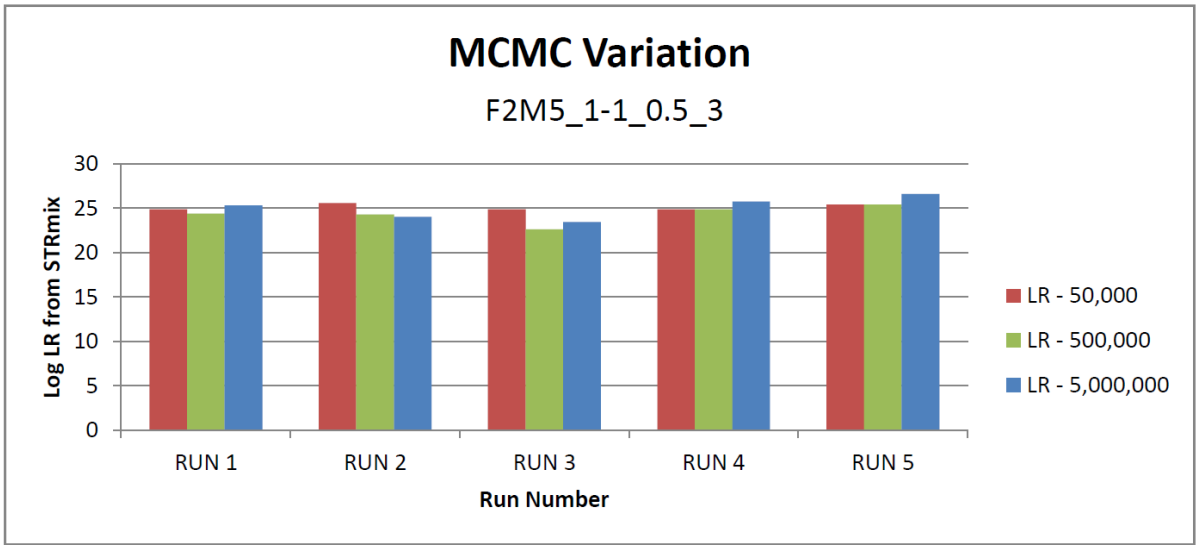


Figure M2. Plot of log(LR) versus MCMC iterations for sample F2M5_1-1_0.5_3 in the Palm Beach Fusion African American population data

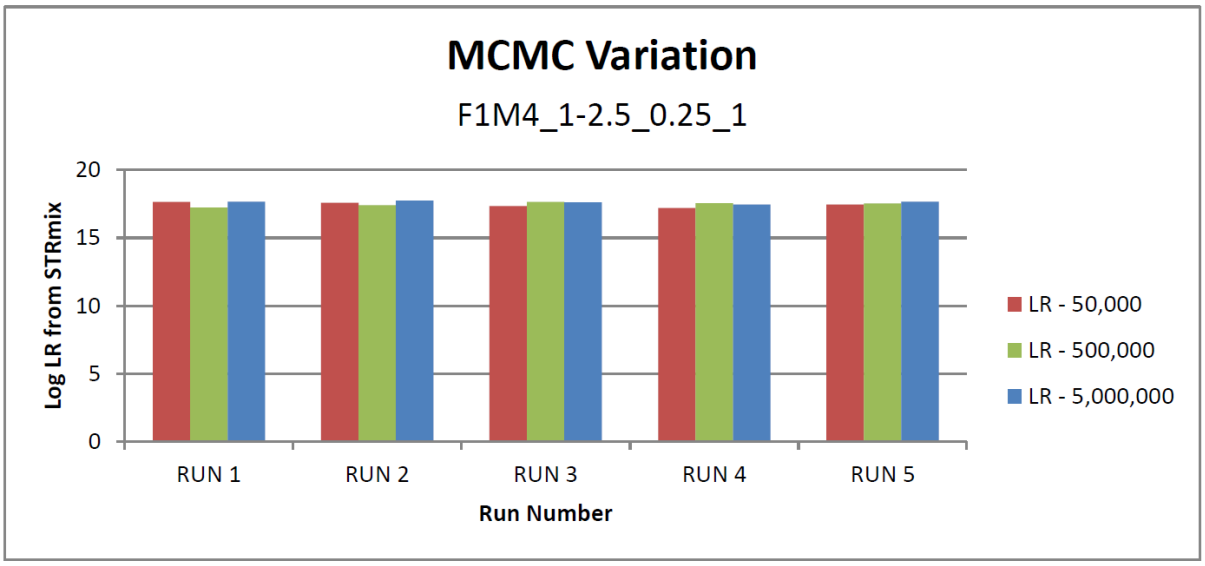


Figure M3. Plot of log(LR) versus MCMC iterations for sample F1M4_1-2.5_0.25_1 in the Palm Beach Fusion Hispanic population data.

Conclusion

The internal validation for STRmix™ V2.4has shown that STRmix™ is suited for its intended use for the interpretation of profiles generated from casework samples.

The following summary lists the limitations of the software and recommendations for use in casework analysis. This section is designated to help provide documentation for the connection of the validation to standard operating procedures.

Section A: Single source profiles

Likelihood ratios for known contributors decrease with lower templates. Therefore, STRmix™ is giving more weight to possible genotypes that include dropout. Additionally, the likelihood ratios calculated by STRmix™ were verified with the published formulas using Microsoft Excel.

Section B: Use of peak heights

If a single source sample is saturated (peaks above 30,000 RFU), STRmix™ will correctly interpret the profile qualitatively.

LIMITATION: Because STRmix™ uses stutter and allele peak heights, mixed samples will not be properly interpreted if peak heights are saturated.

Recommendation: Where practicable, mixed samples which are saturated should be reamplified at a lower template and re-run.

Section C: Weights

STRmix™ appropriately decreased the likelihood ratio/weights as the ability to define a contributor decreased. Where the major contributor is in high template (500 pg) and has a high mixture proportion, the LR obtained approaches that obtained from the contributors single source profiles as the genotypes of the major contributor can be clearly resolved. When the mixture proportions move towards a 1:1 mixture ratio there are more genotype combinations possible to explain the observed profile and therefore the weights for the known major contributor's genotype combinations decrease as do the LRs obtained. For the minor contributor, at low mixture proportions, the template amount of DNA for this contributor is low and some alleles may not be distinguished above the laboratory's analytical threshold or considered as possible stutter. An increase in mixture proportions and template DNA may allow more alleles to be both detected and distinguished from the major contributor thus causing the LR to rise. At mixture proportions close to 1:1 more alleles may be detected for the minor contributor, however, the LR may decrease due to ambiguity in assigning genotypes to individual contributors.

Section D: Sensitivity, specificity, and mixtures

Sensitivity- The ability of the software to reliably resolve the DNA profile of known contributors within a mixed DNA profile for a range of starting DNA templates.

Specificity- The ability of the software to reliably exclude known non-contributors within a mixed DNA profile for a range of starting DNA templates.

LIMITATION: Based on the samples run, false inclusions and exclusions can occur with low level contributors in two, three, and four person mixtures. Regardless of the template quantity, no false inclusions or exclusions were observed for single source samples. Because validation samples are specifically chosen to create mixtures with varying alleles, it is expected for casework samples to show a slightly larger range of false inclusions and exclusions.

The data obtained during the Model Maker work up shows that there is high allele and stutter variance values for PowerPlex® Fusion at 30 cycles. High variance values will make the acceptance or rejection criteria for proposed genotype combinations during MCMC more tolerant of differences between observed and expected profiles (accepting more combinations) but close matches between the observed and the expected will be less well rewarded than they would be under lower variance values. Therefore, there is likely to be a wider zone of LRs that might be considered "uninformative" and this 'zone' may come into play at higher values of template/APH than with platforms with lower variance.

Recommendation: Where profile peak heights or templates are low and for four person mixtures it may be beneficial to run replicate amplifications and/or increased iterations (x10 more iterations).

Section E: Alternate propositions

Assuming a known contributor can improve the performance of the software especially at lower template amounts. It is critical that any assumptions made while interpreting the data are clearly documented.

Section F: Assigning the number of contributors

Over estimating the number of contributors can reduce the LR of true contributors, notably for minor, low level contributors. Increasing the number of contributors can increase the number of positive LR's for known non contributors (false inclusion).

Subtracting a contributor had little effect on the LR for some known contributors (the major/intermediate components) but a false exclusion for some (the minor/trace components) contributors were obtained. This is not unexpected as STRmix™ has to spread the genotype combinations across fewer proposed contributors and hence some genotype combinations of the true contributors are not considered.

LIMITATION: The true number of contributors is never known for evidence profiles. Incorrectly assuming the number of contributors which is above or below the correct number of contributors may lead to slightly lower likelihood ratios and false inclusions/exclusions (LR above one/LR below one).

Section G: Drop-in

When considering drop-in, LRs will not change when the drop-in peak is within model parameters and considered as drop-in by STRmix™.

LIMITATION: One sample, C03_F1M4_5-1_2 , from section D contained a small peak present at the '10' allele position at Penta D. The "10" peak was beneath the drop-in cap established during Model Maker. Given the profile data, STRmix ™ may consider the "10"as possible 'drop-in' or truly allelic. If considered as 'drop-in' then positive Log(LR) values would obtained when compared to the reference profile. If the '10' peak is considered as allelic then the contributor may be excluded. This is a run to run variability issue.

Recommendation:

Reviewing the primary diagnostics (i.e., mixture proportions, weights, LR per locus, log likelihood, allele and stutter variance, Gelman Rubin Score, etc.) in the advance report would help to identify possible run to run variability and to evaluate the results to make sure they are intuitively correct. If one locus produces a significantly different result than all the other loci, the profile should be re-evaluated.

Section H: Forward and reverse stutter

STRmix™ correctly models forward and reverse stutter. A true stutter peak above the expected value may produce a false exclusion and true alleles eliminated due to its presence in a stutter position may result in either a reduced likelihood ratio or exclusion.

Section I: Intra-Locus peak height

Model Maker results were verified with regard to heterozygote balance and approximate peak height. As expected, heterozygote balance is more variable at lower peak heights.

LIMITATION: The data demonstrate a high variability in peak heights for both alleles and stutter peaks. This may not necessarily be unexpected given the combination of the kit, cycle number and CE platform used. There is likely to be a wider zone of LRs that might be considered "uninformative" and this 'zone' may come into play at higher values of template/APH than with platforms with lower variance.

Section J: Inter-Locus peak height

Model maker results were verified for locus specific amplification efficiencies (LSAE). Average peak heights were compared to LSAE results to demonstrate concordance.

Section K: Challenge testing

STRmix™ requires that only numeric values are retained within the input file. Samples with non-stutter artifacts may produce one of three results; exclusionary LR, no effect, or failure of the software to interpret the data. All of the artifacts added to the input file contained artifacts that were easily identified but could be missed during analysis. During validation, it was determined that most of the time, software errors are due to users errors during analysis.

LIMITATION: Artifacts or peaks in stutter position should be closely evaluated and carefully eliminated. Errors present in the.txt file may result in STRmix™ errors or false exclusions.

Recommendation: During review of the STRmix™ results, weights and likelihood ratios should be carefully evaluated to make sure they are intuitively correct. If one locus produces a significantly different result than all others, re-evaluate the locus to determine if an artifiact may be an issue.

Section L: Non-Probative samples

Eleven non-probative casework samples were interpreted by STRmix™. Concordant results were obtained between the manual interpretation and the LR produced with STRmix™ with the exception of two separate sample/reference combination for NP 11 (it should be noted that NP11 is a four person mixture). When compared to 11B the reference samples was interpreted as inconclusive but returned a LR of 2.33E+13. When compared to 11C, the reference sample was interpreted as inconclusive but returned a LR of 1.2E-15. Given the allele and stutter variance of the kit there is likely to be a wider zone of LRs that might be considered "uninformative".

Section M: Precision

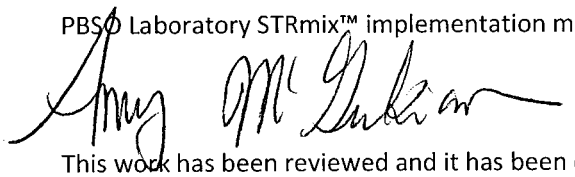
As expected, STRmix™ did not give the same LR each time, however, the variability was low. Additionally, the highest posterior density (HPD) was always lower than the likelihood ratio demonstrating its ability to account for sampling variation in the allele frequency database. Lower iterations produce more variability than higher iterations, but higher iterations may have longer run times.

LIMITATION: Variability in the LR is not only affected by the complexity of the profile, but also the chosen number of iterations.

Recommendation: Run all samples at the default setting of 500,000 total iterations (100,000 burn-ins). This will result in the least amount of variability with a reasonable run time. Raising the iterations may be useful for highly complex profiles to reduce variability or for trouble shooting of samples. Raising the number of iterations to 5,000,000 (x10) may be useful in these instances.

Signatures

PBSO Laboratory STRmix™ implementation manager



This work has been reviewed and it has been determined that STRmix™ V2.4 is suitable for its intended use for interpretation of crime profiles at PBSO Laboratory. The project work has met the validation requirements as required by SWGDAM.

Amy McGuckian, PBSO Laboratory Technical Leader

References

[1] Taylor D, Bright J-A, Buckleton JS. The interpretation of single source and mixed DNA profiles Forensic Science International: Genetics. 2013;7:516-28.

[2] Scientific Working Group on DNA Analysis Methods (SWGDAM). Guidelines for the Validation of Probabilistic Genotyping Systems. 2015.

[3] Bright J-A, Evett IW, Taylor D, Curran JM, Buckleton J. A series of recommended tests when validating probabilistic DNA profile interpretation software. Forensic Science International: Genetics. 2015;14:125-31.

[4] Balding DJ, Nichols RA. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Science International. 1994;64:125-40.

[5] National Research Council II. National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence. Washington, D.C.: National Academy Press; 1996.

[6] Taylor D. Using continuous DNA interpretation methods to revisit likelihood ratio behaviour. Forensic Science International: Genetics. 2014;11:144-53.

[7] Taylor D, Bright J-A, Buckleton J. The 'factor of two' issue in mixed DNA profiles. Journal of Theoretical Biology. 2014;363:300-6.

[8] Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US Caucasians, Hispanics, Bahamanians, Jamaicans and Trinidadians. Journal of Forensic Sciences. 1999;44:1277-86.

[9] Bright J-A, Stevenson KE, Curran JM, Buckleton JS. The variability in likelihood ratios due to different mechanisms. Forensic Science International: Genetics. 2015;14:187-90.

[10] Taylor D, Bright J-A, Buckleton J, Curran J. An illustration of the effect of various sources of uncertainty on DNA likelihood ratio calculations. Forensic Science International: Genetics. 2014;11:56-63.

[11] Triggs CM, Curran JM. The sensitivity of the Bayesian HPD method to the choice of prior. Science & Justice. 2006;46:169-78.

[12] Curran JM, Buckleton JS. An investigation into the performance of methods for adjusting for sampling uncertainty in DNA likelihood ratio calculations. Forensic Science International: Genetics. 2011;5:512-6.

APPENDIX 1: List of papers that support STRmix™

The following is a list of papers that directly support STRmix™.

1. D. Taylor, J.-A. Bright and J.S. Buckleton, The interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics, 2013 7(5): 516-528 **(Core maths paper)**
2. J.-A. Bright, D. Taylor, J.M. Curran and J.S. Buckleton, Developing allelic and stutter peak height models for a continuous method of DNA interpretation. Forensic Science International: Genetics, 2013. 7(2): 296-304 **(Core models paper)**
3. J.-A. Bright, D. Taylor, J.M. Curran and J.S. Buckleton, Degradation of forensic DNA profiles, Australian Journal of Forensic Sciences, 2013. 45(4): 445-449
4. D. Taylor. Using continuous DNA interpretation methods to revisit likelihood ratio behavior. Forensic Science International: Genetics, 2014. 11: 144-153
5. J.-A. Bright, D. Taylor, J.M. Curran and J.S. Buckleton, Searching mixed DNA profiles directly against profile databases. Forensic Science International: Genetics, 2014. 9: 102-110
6. D. Taylor, J.-A. Bright, J.S. Buckleton, J. Curran, An illustration of the effect of various sources of uncertainty on DNA likelihood ratio calculations. Forensic Science International: Genetics, 2014. 11: 56–63
7. J.-A. Bright, J.M. Curran and J.S. Buckleton, The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation. Forensic Science International: Genetics, 2014. 12: 208-214
8. J.-A. Bright, K.E. Stevenson, J.M. Curran and J.S. Buckleton, The variability in likelihood ratios due to different mechanisms. Forensic Science International: Genetics, 2015. 14:187-190
9. D .Taylor, J.-A. Bright and J.S. Buckleton, Considering relatives when assessing the evidential strength of mixed DNA profiles. Forensic Science International: Genetics, 2014. 13: 259-263
10. D. Taylor, J-A. Bright and J.S. Buckleton. Interpreting forensic DNA profiling evidence without specifying the number of contributors. Forensic Science International: Genetics, 2014. 13: 269-280

The following is a subset of other papers that support the theory within STRmix™:

1. J.-A. Bright, J.M. Curran. Investigation into stutter ratio variability between different laboratories. Forensic Science International: Genetics, 2014. 13: 79-81
2. C. Brookes, J.-A. Bright, S.A. Harbison, and J.S. Buckleton, Characterising stutter in forensic STR multiplexes. Forensic Science International: Genetics, 2012. 6(1): 58-63
3. H. Kelly, J.-A. Bright, J.M. Curran, and J.S. Buckleton Identifying and modelling the drivers of stutter in forensic DNA profiles. Australian Journal of Forensic Sciences, 2014. 46(2): 194-203
4. J.-A. Bright, S. Neville, J.M. Curran, and J.S. Buckleton. Variability of mixed DNA profiles separated on a 3130 and 3500 capillary electrophoresis instrument. Australian Journal of Forensic Sciences, 2014. 46(3): 304-312
5. J.-A. Bright, K.E. Stevenson, M.D. Coble, C.R. Hill, J.M. Curran, and J.S. Buckleton Bright, Characterising the STR locus D6S1043 and examination of its effect on stutter rates. Forensic Science International: Genetics, 2014. 8(1): p. 20-23.
6. D. Taylor, J.S. Buckleton. Do low template DNA profiles have useful quantitative data? Forensic Science International: Genetics, 2015. 16: 13-16.

The following is a subset of other papers that support the validation and use of STRmix™:

1. J.-A. Bright, I.W. Evett, D. Taylor, J.M. Curran and J.S. Buckleton, A series of recommended tests when validating probabilistic DNA profile interpretation software. Forensic Science International: Genetics, 2015. 14: 125-131
2. T.W. Bille, S.M. Weitz, M.D. Coble, J.S. Buckleton, J.-A. Bright. Comparison of the performance of different models for the interpretation of low level mixed DNA profiles. ELECTROPHORESIS. 2014;35:3125-33.
3. S.J. Cooper, C.E. McGovern, J.-A. Bright, D. Taylor, J.S. Buckleton. Investigating a common approach to DNA profile interpretation using probabilistic software. Forensic Science International: Genetics, 2014. 16: 121-131.

Appendix 2: Cross reference for document sections and SWGDAM recommendations

Standard	Text	Refer section
4.1	Test the system using representative data	Preamble
4.1.1	Specimens with known contributors	Preamble
4.1.2	Hypothesis testing with contributors and non-contributors	D
4.1.2.1	More than one set of hypotheses	E
4.1.3	Variable DNA typing conditions	Preamble
4.1.4	Allelic peak height, to include off-scale peaks	B
4.1.5	Single-source specimens	A
4.1.6	Mixed specimens	D
4.1.6.1	Various contributor ratios	D
4.1.6.2	Various total DNA template quantities	D
4.1.6.3	Various numbers of contributors	D
4.1.6.4	Both correct and incorrect number of contributors (i.e., over- and under-estimating)	F
4.1.6.5	Sharing of alleles among contributors	D
4.1.7	Partial profiles	D
4.1.7.1	Allele and locus drop-out	D
4.1.7.2	DNA degradation	L
4.1.7.3	Inhibition	L
4.1.8	Allele drop-in	G
4.1.9	Forward and reverse stutter	H
4.1.10	Intra-locus peak height variance	I
4.1.11	Inter-locus peak height variance	J
4.1.12	In-house parameters	Preamble
4.1.13	Sensitivity, specificity and precision	D and M
4.1.14	Additional challenge testing	K
4.2	Compare the results of probabilistic genotyping and of manual interpretation	L
4.2.1	Intuitive and consistent with expectations	L
4.2.1.1	Known specimens that are included based on non-probabilistic analyses would be expected to also be included based on probabilistic genotyping	L
4.2.1.2	Concordance of single-source specimens with high quality results	A
4.2.1.3	Generally, as the analyst's ability to deconvolute a complex mixture decreases, so does the weighting of a genotype set determined by the software	C