STRmix™
RESOLVE MORE DNA MIXTURES.

Palm Beach County

Sheriff's Office

Internal Validation of STRmix™ V2.6.2 (PowerPlex® Fusion6C, 3500xlCE)

REVIEWED
Date: 10/29/19
Initials: ASM
OK for use on casework

STRmix™ internal validation

This document describes the internal validation of STRmix™ V2.6.2 for PowerPlex® Fusion 6C profiles analysed on a 3500xl Genetic Analyzer within the Palm Beach County Sheriff's Office system (hereafter, PBSO).

Note that this is a validation of STRmix™ V2.6 and not the PowerPlex® Fusion 6C chemistry or 3500xl capillary electrophoresis (CE) technology as these were validated separately (See Internal Validation for PowerPlex Fusion 6C® on the 3500xl). STRmix™ V2.4.08 has previously been internally validated at PBSO for PowerPlex® Fusion 5C profiles analysed on a 3500xl CE platform. This validation was done in conjunction with the STRmix™ Scientific Support team of ESR in New Zealand.

STRmix™ has previously been subjected to developmental validation. This included specificity and sensitivity mixture studies. 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 Bright et al. (1). A list of all papers describing the theory behind different aspects of STRmix™ is provided in Appendix 1 of this document.

This validation of STRmix™ V2.6 for PowerPlex® Fusion 6C profiles is only a *subset* of the tasks undertaken in the original work and follows elements of the internal validation section of the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems [1]. Parts of the validation that were previous covered during the validation of STRmix v 2.4 and are not impacted by the STR typing chemistry or electrophoresis platform used to collect the STR typing data were not conducted as part of the STRmix™ v 2.6 validation. This internal validation follows the approach recommended by the STRmix™ developers in the report "STRmix™ V2.6.0 Release and Testing Report", 15 August 2018.

The section where specific SWGDAM guidelines are discussed in this document is cross referenced in Appendix 2.

The samples used within this validation and described in this report were all generated at the PBSO laboratory.

The results of all experiments related to the internal validation of STRmix™ at PBSO may be found in the associated companion binder (electronic binder).

STRmix™ parameters

The parameters described in the document "Estimation of STRmix™ v2.6 parameters for the Palm Beach County Sheriff's Office (PowerPlex™ Fusion 6C, 3500xl CE)" 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.

Within this section it is demonstrated that the weights assigned by STRmix™ to different genotype combinations are appropriate. The weights can be used as a diagnostic of the deconvolution process and should be intuitively correct (meet quality control expectations), where the most supported genotypes have the highest weights.

The addition of information to an analysis can aid in the ability to deconvolute the sample. For example, using replicates can reduce ambiguity and increase weightings of individual genotype sets.

Dilution series of two single source samples (#72, #73) were constructed, where the resultant peak heights ranged from above the level where dropout is observed to below. These samples were amplified using Fusion 6C™ following PBSO's validated method. The template DNA amounts in nanograms for the serial dilutions studied here were: 0.4, 0.2, 0.1, 0.05, 0.025 and 0.0125ng. Full single source profiles were obtained at 0.2ng and above. The profiles were analysed using PBSO's validated analytical thresholds.

The profiles were interpreted in STRmix™ and an *LR* calculated considering the known donor as the person of interest. The propositions were:

> H_p: The DNA originated from the person of interest
>
> H_d: The DNA originated from an unknown individual

The Likelihood Ratio (*LR)* was calculated for the known contributor using the NIST 1036 Caucasian allele frequencies and an F_{ST} (θ) of 1%. In relation to the STRmix™ output the point estimate value (Sub Source *LR*) has been selected for direct comparison between runs. Where more than one replicate at each dilution has been used the average log(*LR*) has been plotted. A plot of log(*LR*) versus input DNA for Fusion 6C™ is provided for each sample in Figure 1a and 1b.

Figure 1a: Plot of log(*LR*) versus input DNA amount (pg) for a dilution series of #72 DNA

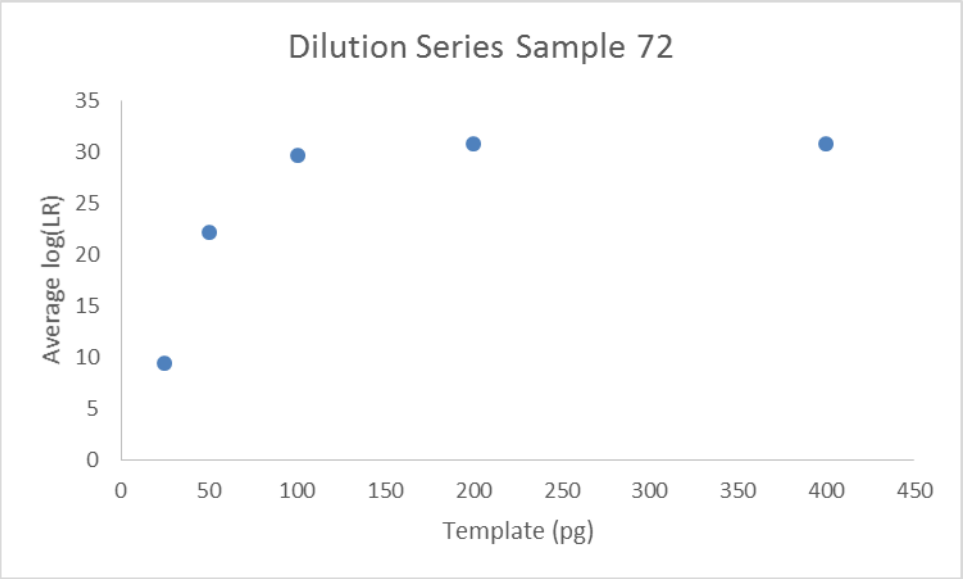
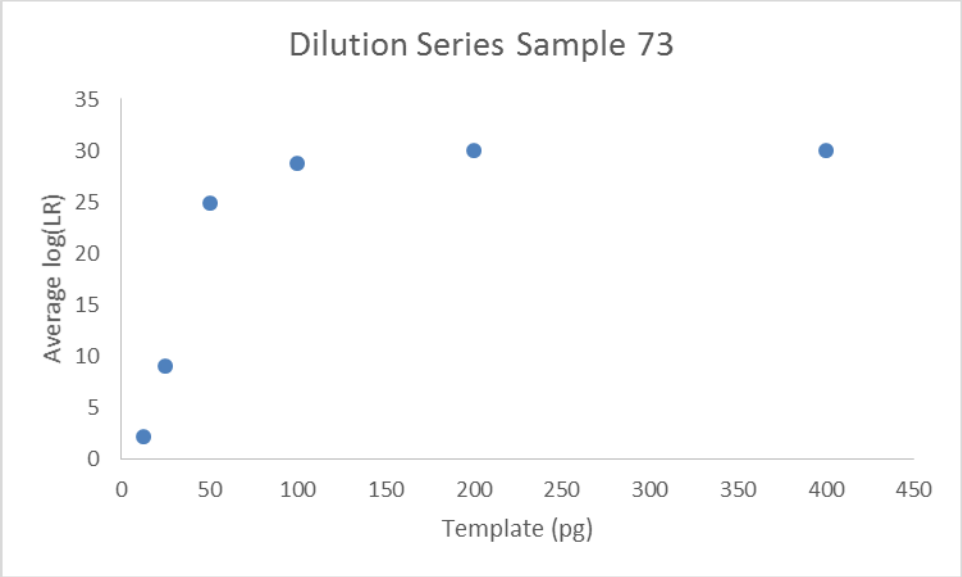


Figure 1b: Plot of log(*LR*) versus input DNA amount (pg) for a dilution series of #73 DNA



Inspection of

Figure 1a and 1b show the *LR* progressing from the value for the single source *LR* calculated for a full profile towards an *LR* = 1 (Log(*LR*) = 0) as the DNA template decreases. As expected, the weights for genotypes considering dropout increased as template drops. In addition, the DNA amounts from the STRmix™ output (*t* or template mass parameter) declined steadily in line with peak heights.

Reproduction of single source LR

There is a small subset of profiles where the 'answer' is known or can be estimated easily (3). These include single source profiles where the weight is one (or 100%) for a single genotype set at each locus. The *LR* was calculated 'by hand' at each locus for two single source profiles (# 72 and # 73) and the individual locus *LR*s compared with the STRmix™ results. This was undertaken twice; once using an F_{ST} (or θ) value of 0 and once with F_{ST}=0.01. Setting θ to zero returns the product rule where:

$2p_ip_j$ for heterozygote loci

p_i^2 for homozygote loci

Where p_i is the allele frequency for allele *i*, p_j the allele frequency for allele *j*. When θ > 0, the Balding and Nichols (4) formulae (or equations 4.10 from NRCII (5)) are applied. For single source profiles:

$$\frac{2\left[\theta+(1-\theta)p_i\right]\left[\theta+(1-\theta)p_j\right]}{(1+\theta)(1+2\theta)} \quad \text{for heterozygote loci} \qquad [1]$$

$$\frac{\left[3\theta+(1-\theta)p_i\right]\left[2\theta+(1-\theta)p_i\right]}{(1+\theta)(1+2\theta)} \quad \text{for homozygote loci} \qquad [2]$$

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. The allele frequencies used within equations 1 and 2 are posterior mean frequencies. These are calculated using the following equation:

$$\frac{x_i+\frac{1}{k}}{N_a+1} \qquad [3]$$

Where for the given locus, xi 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 at that locus.

The 'by hand' calculated and STRmix™ results for the single source profiles for θ=0 and θ=0.01 are given in Tables 1a and 1b.

Table 1a: 'By hand' (Excel) calculation of *LR* versus STRmix™ results for a full single source profile (# 72) with varying F_{ST} values. CSF1PO (θ =0) and TPOX (θ=0.01) differences between Excel and STRmix™ are due to rounding.

Locus	Excel θ=0	STRmix™ θ=0	Excel θ=0.01	STRmix™ θ=0.01
D3S1358	9.9834	9.9834	9.6054	9.6054
D1S1656	46.7028	46.7028	40.5042	40.5043
D2S441	6.9268	6.9268	6.7486	6.7486
D10S1248	12.2434	12.2434	11.5797	11.5797
D13S317	39.4541	39.4541	32.9218	32.9218
Penta E	32.9570	32.9570	29.1087	29.1087
D16S539	27.9675	27.9675	24.1879	24.1879
D18S51	162.5329	162.5329	103.4514	103.4514
D2S1338	41.3469	41.3469	36.4502	36.4502
CSF1PO	20.6396	20.6397	17.4664	17.4664
Penta D	9.7143	9.7143	9.3587	9.3587
TH01	12.1815	12.1815	11.4671	11.4671
vWA	90.2625	90.2625	61.8052	61.8052
D21S11	18.0237	18.0237	16.5859	16.5859
D7S820	15.3300	15.3300	14.4403	14.4403
D5S818	9.0492	9.0492	8.6570	8.6570
TPOX	39.7255	39.7255	33.3951	33.3952
D8S1179	9.0645	9.0645	8.7185	8.7185
D12S391	42.0204	42.0204	36.9484	36.9484
D19S433	5.4390	5.4390	5.3492	5.3492
SE33	620.4588	620.4588	247.2359	247.2359
D22S1045	9.3601	9.3601	8.9395	8.9395
FGA	106.1777	106.1777	76.3803	76.3803
Total	**2.7031E+32**	**2.7031E+32**	**6.2010E+30**	**6.2010E+30**

Table 1b: 'By hand' (Excel) calculation of *LR* versus STRmix™ results for a full single source profile (# 73) with varying F_{ST} values. D12S391 (θ =0) difference between Excel and STRmix™ due to rounding.

Locus	Excel θ=0	STRmix™ θ=0	Excel θ=0.01	STRmix™ θ=0.01
D3S1358	87.9109	87.9109	60.4997	60.4997
D1S1656	123.3935	123.3935	96.5870	96.5870
D2S441	23.8939	23.8939	20.9303	20.9303
D10S1248	8.5498	8.5498	8.2669	8.2669
D13S317	5.7262	5.7262	5.6259	5.6259
Penta E	38.8975	38.8975	34.0594	34.0594
D16S539	5.0658	5.0658	4.9982	4.9982
D18S51	35.7573	35.7573	31.8983	31.8983
D2S1338	56.5397	56.5397	48.2003	48.2003
CSF1PO	7.3600	7.3600	7.1624	7.1624
Penta D	17.9271	17.9271	16.6835	16.6835
TH01	18.0609	18.0609	15.4885	15.4885
vWA	24.8280	24.8280	20.6000	20.6000
D21S11	15.5516	15.5516	14.6390	14.6390
D7S820	11.6582	11.6582	11.1185	11.1185
D5S818	7.9074	7.9074	7.2473	7.2473
TPOX	3.7853	3.7853	3.7531	3.7531
D8S1179	14.6166	14.6166	13.5855	13.5855
D12S391	208.5698	208.5699	118.0520	118.0520
D19S433	19.6046	19.6046	17.5379	17.5379
SE33	185.8888	185.8888	134.8889	134.8889
D22S1045	11.1344	11.1344	10.5823	10.5823
FGA	124.5056	124.5056	88.7443	88.7443
Total	**2.4998E+31**	**2.4998E+31**	**8.9539E+29**	**8.9539E+29**

The results in Tables 1a and 1b 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. This can be observed in

Figure 1 above (and later in Figure 2). This is the expected result.

It is not recommended that saturated mixture profiles are interpreted in STRmix™ as a profile that exceeds the CE saturation threshold is unlikely to have the true peak heights recorded. Thus the models used within STRmix™ are no longer optimal and it is likely that higher stutter peak heights than expected will be observed relative to the allele heights recorded in the electropherograms, resulting in an elevated k^2 value.

A saturation setting of 30,000 rfu was selected for PBSO in the previous v2.4 validation. In the validation for STRmix V2.4, 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.

DNA mixture profiles with peaks exceeding 30,000 rfu will not be used for DNA mixture interpretation with STRmix™.

Section C: Mixture 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.

This principle was demonstrated in the previous validation of V2.4. The results obtained from the sensitivity study in Section D, described below, also show the weights obtained from the deconvoluted mixtures are generally intuitive.

Section D: Sensitivity and 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's Fusion 6C™ mixtures was undertaken based on Taylor (8), however making use of average peak heights, rather than template. 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 0 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) and replicates. A downward trend in the *LR* is also expected with increasing numbers of contributors as the complexity of the mixture is compounded. Specificity is defined as 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 0 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. Table 2 summarises the mixed DNA samples prepared for this study.

Table 2: Experimental design for sensitivity and specificity testing

Two Person Mixtures (Mix 04)

Ratio	Total DNA amount (ng)
20:1	1, 0.5, 0.25, 0.1
10:1	1, 0.5, 0.25, 0.1
5:1	1, 0.5, 0.25, 0.1
2:1	1, 0.5, 0.25, 0.1
1:2	1, 0.5, 0.25, 0.1
1:5	1, 0.5, 0.25, 0.1
1:10	1, 0.5, 0.25, 0.1
1:20	1, 0.5, 0.25, 0.1
Contributor 1	M3
Contributor 2	F8

Three Person Mixtures (Mix 06)

Ratio	Total DNA amount (ng)
10:5:1	1, 0.5, 0.25, 0.1
8:1:1	1, 0.5, 0.25, 0.1
3:2:1	1, 0.5, 0.25, 0.1
1:1:1	1, 0.5, 0.25, 0.1
Contributor 1	M24
Contributor 2	F3
Contributor 3	F9

Four Person Mixtures (Mix 08)

Ratio	Total DNA amount (ng)
10:5:2:1	1, 0.5, 0.25, 0.1
9:3:3:1	1, 0.5, 0.25, 0.1
6:3:1:1	1, 0.5, 0.25, 0.1
4:4:1:1	1, 0.5, 0.25, 0.1
4:3:2:1	1, 0.5, 0.25, 0.1
1:1:1:1	1, 0.5, 0.25, 0.1
Contributor 1	M28
Contributor 2	M36
Contributor 3	F19
Contributor 4	F24

The plots in (8) have been reproduced for PBSO's Fusion 6C data. Each mixed sample was prepared in duplicate and amplified according to PBSO's validated method. The replicate amplifications were differentiated by the suffix of '**a**' or '**b**'. 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). These profiles include some of the 'worst' types of profiles likely to be encountered by the laboratory.

Analysis of the mixed DNA profiles was carried out at the PBSO's validated analytical thresholds of:

Blue: 75 rfu, Green: 101 rfu, Yellow: 60 rfu, Red: 69 rfu, Purple: 56 rfu, Orange: 50rfu.

Each profile was interpreted in STRmix™ using the experimental design number of contributors and compared to the known contributors and 490+ known non-contributors using the Database Search function within STRmix™. The non-contributors were generated by PBSO blood standards and buccal swabs from known non-contributors. The non-contributor samples were typed with PowerPlex Fusion 6C. There were a total of nine known contributors across all the mixtures and 493 non-contributors created. Therefore, for any given mixture there would be approximately 500 non-contributors compared to each mixture output.

Using the NIST Revised Caucasian Allele Frequencies and an F_{ST} of 0.01, an *LR* was calculated where the following propositions considered were:

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

H_d: The DNA originated from *N* unknown individuals

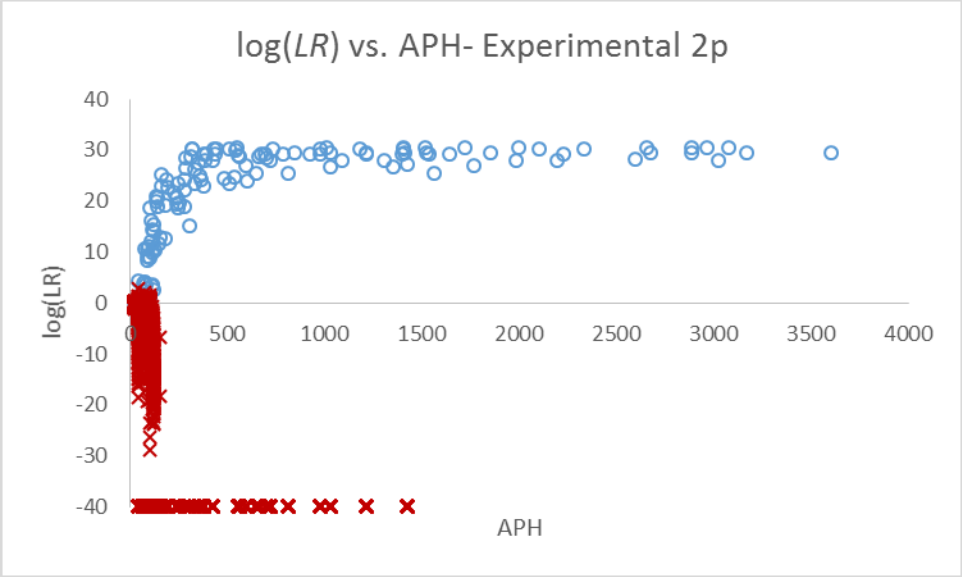
Where *N* is the experimental design number of contributors to the profile.

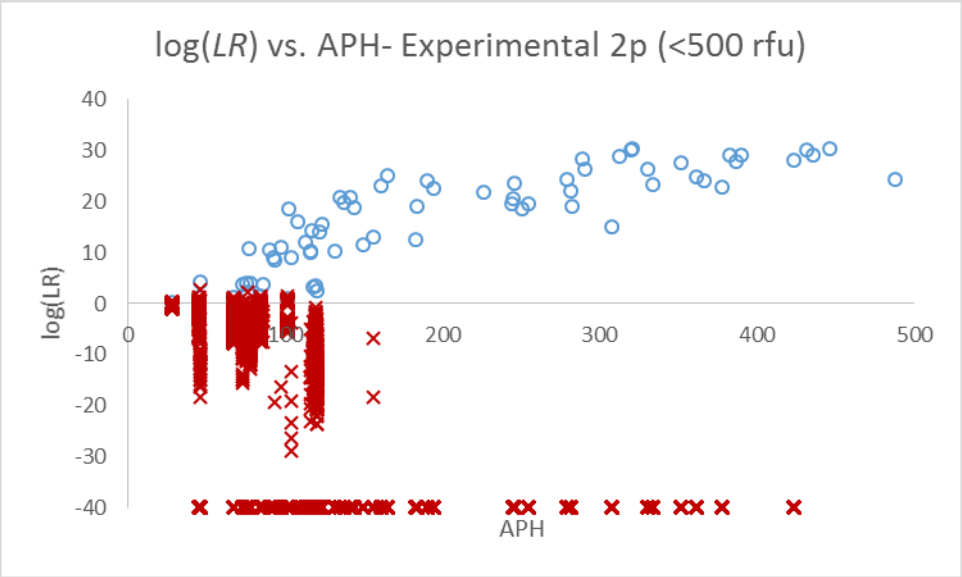
Plots of log(*LR*) versus the average peak height (*APH*) per known contributor for the two-, three-, and four-contributor mixtures are given in Figure 2. *APH* was calculated using unmasked, unshared, and non-stutter affected alleles for each known contributor in the mixed profiles. Where the contributor had completely dropped out of the mixture, an *APH* of half the analytical threshold (AT) used by PBSO is applied. The per-contributor amount of DNA for known non-contributors is taken as the lowest APH of the known contributors per mixture, which may also be the half AT value in some instances.

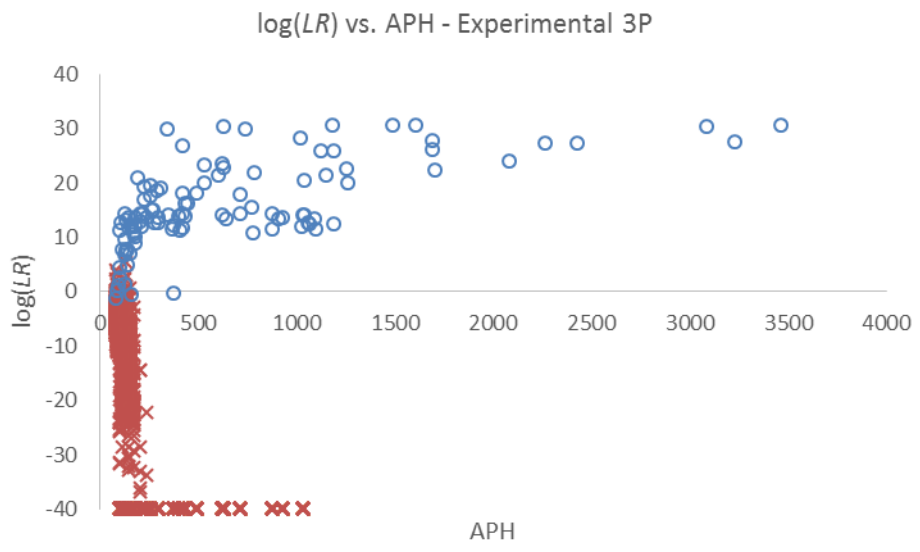
Exclusions (*LR* = 0) for known contributors and non-contributors were plotted as log(*LR*) = -40. The results of all comparisons are provided in Figure 2, where the blue circles indicate a known contributor to the mixture (H_p true), and red crosses indicate a known non-contributor (H_d true).

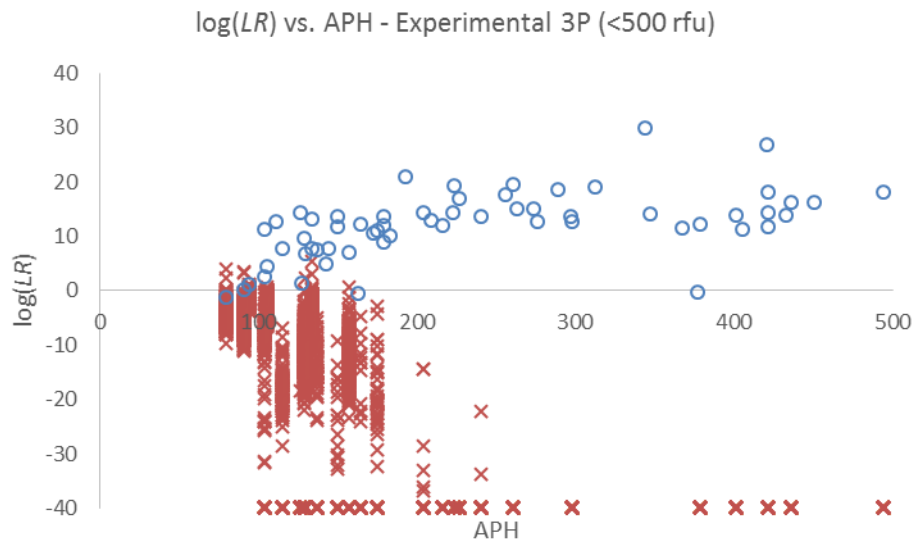
Inspection of the plots in Figure 2 show that as template (and hence *APH*) increases the *LR* distributions for H_p true and H_d true are very well separated for the two-, three- and four-person mixtures. As the number of contributors increases and the template (and hence *APH*) lowers the two distributions converged on *LR*=1 (log(*LR*) = 0). At high template STRmix™ correctly and reliably gave a high *LR* for true contributors and a low (often 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 (*LR*=1).


log(LR) vs. APH- Experimental 2p
log(LR)
APH
40
30
20
10
0
-10
-20
-30
-40
0
500
1000
1500
2000
2500
3000
3500
4000


log(LR) vs. APH- Experimental 2p (<500 rfu)
log(LR)
APH
40
30
20
10
0
-10
-20
-30
-40
0
100
200
300
400
500


log(LR) vs. APH - Experimental 3P
log(LR)
APH
40
30
20
10
0
-10
-20
-30
-40
0
500
1000
1500
2000
2500
3000
3500
4000


log(LR) vs. APH - Experimental 3P (<500 rfu)
log(LR)
APH
40
30
20
10
0
-10
-20
-30
-40
0
100
200
300
400
500

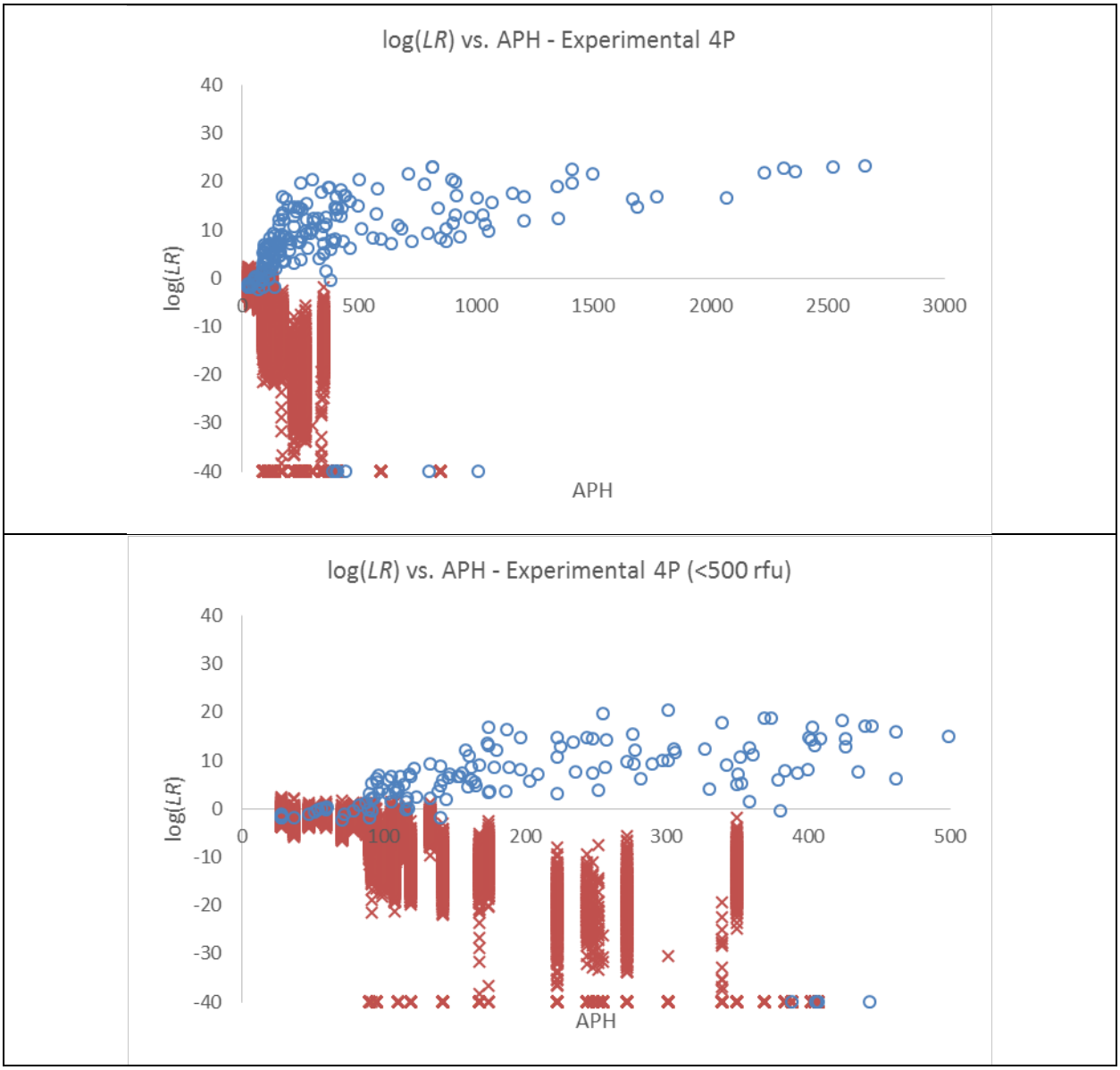
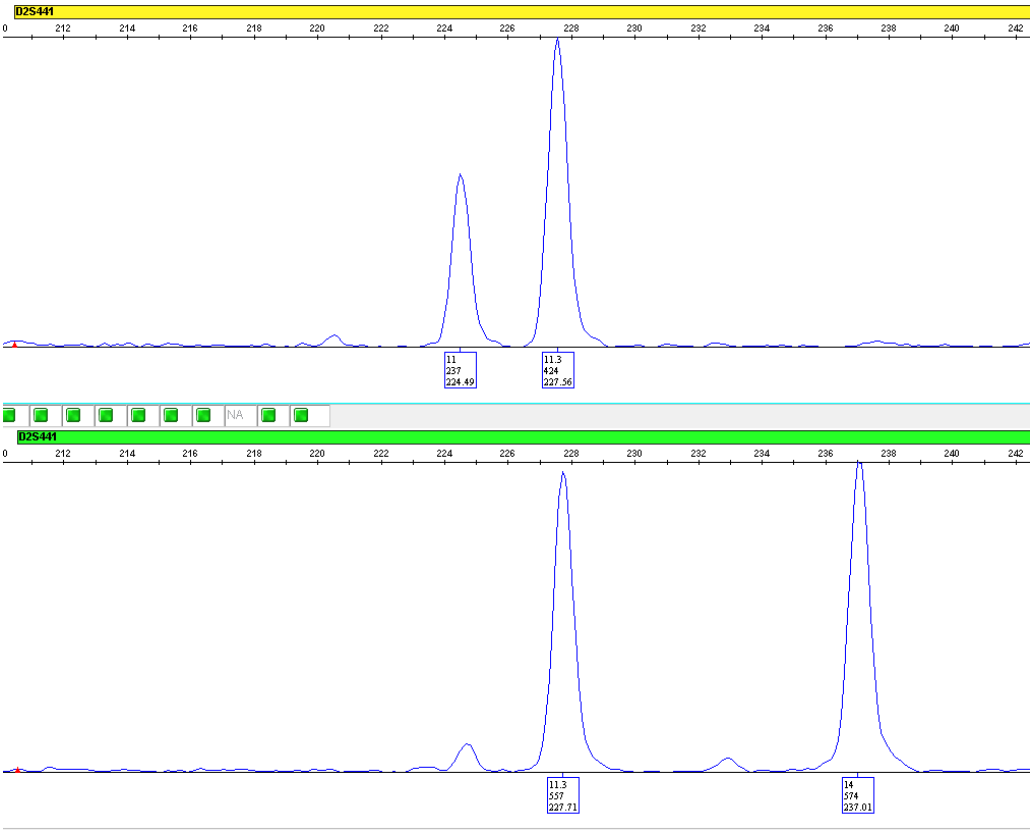


Figure 2: Log(LR) versus APH (rfu) for two-, three- and four-person Fusion 6C mixtures. Every second plot is a close-up to better illustrate the data below an *APH* of 500 rfu.

An *LR* of less than 1 was obtained for the known contributor M24 with an APH value of 377 rfu in one of the three person mixtures, (A11_F_Mix06_0.1_8to1to1_b.hid.csv). M24 was the major contributor to this mixed profile. The *LR* for M24 in the '**a**' replicate of this mixture was 6.77E+29 from a similar APH of 344 rfu. M24 has the genotype 11.3, 14 at D2S441. The very low LR in replicate **b** is most probably due to the drop out of the 14 allele at D2S441. At all other loci M24 corresponds to most highly favored genotype combination for the major contributor. The egram for locus D2S441 (Figure 3) for each of the two replicates is shown below and illustrates why the 11.3, 14 genotype combination was not accepted for the major contributor in replicate **b** but was for replicate **a**. While replicate analysis in STRmix™ is discussed further on, this is a good example of why the analysis of replicate amplifications is beneficial.

Figure 3. Partial Egram (D2S441 only) for replicate amplifications A11_F_Mix06_0.1_8to1to1_b. (top) and D11_F_Mix06_0.1_8to1to1_a. (bottom)



Furthermore, there were exclusions of a known contributor in six mixtures from the four-person mixture set. All relate to individual F19. The exclusion was caused by the lack of resolution of the donor's 15.3 allele from an adjacent 16 allele at locus D1S1656. This 1bp resolution issue is also discussed and illustrated within the discussion around replicates below.

Within STRmix™ the primary diagnostics used to assess the appropriateness of a run are the genotype weights, mixture proportions (Mx) and where undertaken, the individual locus *LR*s. These values should be intuitive relative to the DNA profiles. A review of all the mixture proportions proposed by STRmix™ for these 116 mixtures suggests, on the whole, the Mx's are similar to the intended experimental design.

Secondary diagnostics include the number of iterations, log (likelihood), Gelman-Rubin convergence diagnostic, and posterior mean of the allele and stutter variances. These were generally as expected and a full discussion and summary of these secondary diagnostics for Section D mixture interpretations can be found in Appendix 3.

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.

A sub-set of the profiles used in Section D above were reinterpreted in STRmix™ with alternate propositions.

The nine samples selected are described in Table 3. The selected samples cover a range of template, mixture proportion and complexity.

Table 3: Summary of mixtures re-evaluated with alternate propositions

NOC	Sample	Assumed
2	A05_F_Mix04_0.1_1to2_a.hid	F8
2	E02_F_Mix04_0.5_1to2_a.hid	F8
2	H03_F_Mix04_0.25_1to2_a.hid	F8
3	A11_F_Mix06_0.25_3to2to1_a.hid	M24
3	C10_F_Mix06_0.5_1to1to1_b.hid	M24
3	E04_F_Mix06_1.0_10to5to1_a.hid	M24
4	B05_F_Mix08_0.25_4to4to1to1_a.hid	M28
4	C04_F_Mix08_0.5_6to3to1to1_a.hid	M28
4	D05_F_Mix08_0.25_1to1to1to1_a.hid	M28

In these interpretations one of the contributors was assumed as a known under both H_p and H_d. 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

To mimic typical casework, the major contributor was assumed each time and then likelihood ratios calculated to the remaining potential contributors. Likelihood ratios were calculated as per section D.

A plot of the log(LR) calculated under the different propositions is provided in Figure 4.

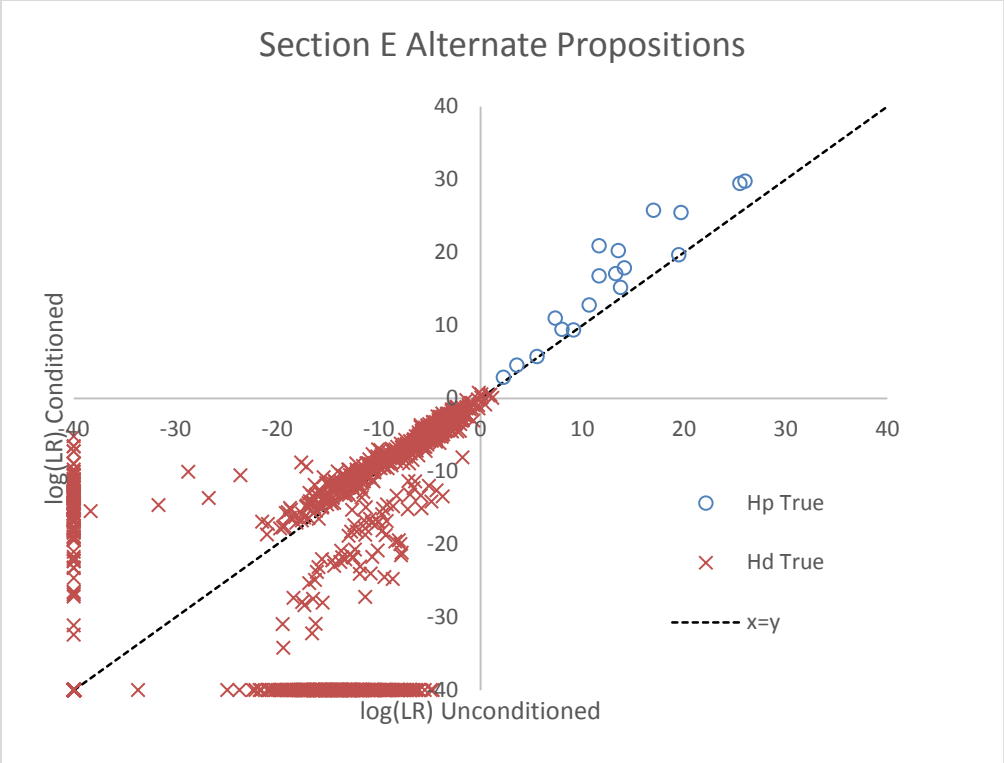


Figure 4: Comparison of the log(*LRs*) obtained from mixtures when assuming a contributor to when there is no person assumed (i.e. section D results). Log(*LR*s) obtained to the known contributors are shown as the blue circles and the non-contributors as red crosses

Values above the dashed line at *x*=*y* for the H_p true *LR*s indicate that the *LR* generally increases when conditioning on, or assuming, a true contributor. Also numerous H_d true *LR*s (red crosses) have also gone from supporting exclusion to now outright exclusion (Log LR -40), when a true contributor is assumed. This shows that the addition of more relevant information (such as the addition of assumed contributors) is shown to improve the performance of STRmix™.

Some of the H_d true *LR*s appear above *x*=*y* line. This is likely due to MCMC variation where a genotype is not 'accepted' in one run but has been 'accepted' in another. Or the relative weight attributed to a genotype combination varies between runs giving altered levels of support for exclusion. This is best illustrated for the four person mixture, sample **C04_F_Mix08_0.5_6to3to1to1_a** where a series of data points track above the x=y trend line. The reduced number of genotype sets given the conditioning of the major contributor has increased the weight of some genotypes for which there is some adventitious correspondence to known non-donors. This has increased their log(*LR*) but these values still very strongly support exclusion.

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.

In casework, the true number of contributors to a questioned 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 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.

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 [9, 10]. 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 comparing the STRmix™ interpretations where the experimental design number of contributors to a mixture (used in section D above) differed by one from the apparent number of contributors to the mixture. There were three mixtures where the apparent number of contributors was one greater than the experimental design. A selection of two and three person mixtures were interpreted as three and four person profiles, respectively (See Tables 4 and 5).

For the purposes of Section F, *N* is defined as the **experimental design number of contributors**. *N*+1 would indicate the addition of one contributor to *N*, and *N*-1 would indicate the subtraction of one contributor to *N*.

Addition of one contributor

During the analysis of the profiles in Section D, three two-person mixtures were assessed as possibly originating from three contributors. Hence these profiles were re-interpreted in STRmix™ as *N*+1. The three samples in question are shown in Table 4.

Table 4: Samples interpreted as originating from *N*+1 contributors

Sample	Design NOC (*N*)	Apparent NOC (*N*+1)
C03_F_Mix04_0.25_10to1_b	2	3
G04_F_Mix04_0.1_1to2_b	2	3
G03_F_Mix04_0.25_2to1_a	2	3

The *LR* for the *N* known contributors and the approximately 500+*N* known non-contributors (as used for the specificity and sensitivity studies, Section D) were calculated from STRmix™ interpretations run assuming *N*+1 contributors. The same allele frequency database and F_{ST} were used as in Section D and the sub source *LR* was used as the point of comparison. The log(*LR*) was compared for the known contributors and known non-contributors under the assumption of *N* and *N*+1 contributors. A plot of log(*LR*) of *N*+1 and *N* is provided in Figure 5.

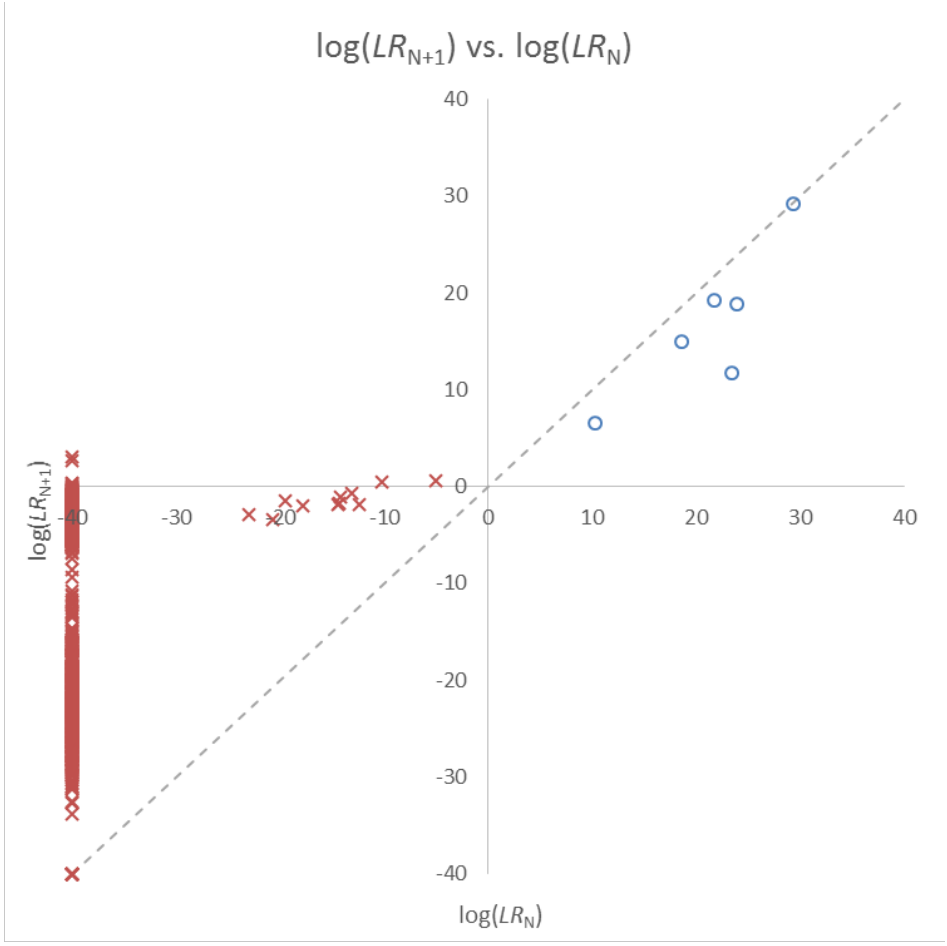


Figure 5: Log(*LR*) when *N* is over-assigned (*N*+1) versus log(*LR*) for experimental design *N* for the known and non-contributors (where H_p true are in blue circles and H_d true in red crosses).

Generally there is a minimal effect on the *LR* of clear major contributors when the number of contributors is over-estimated. One of the mixtures tested here included a major contributor at ratio of 10:1. The log(*LR*) of 30 for this contributor is unchanged by the assumption of an additional contributor. The contributors of the other two mixtures were in 2:1 ratios and the impact of adding an extra contributor is more pronounced. There is a reduction in *LR* in these comparisons, where the weights have been diffused due to the additional contributor being accommodated. This led to the acceptance of more genotype combinations at each locus and the

weightings to be diffused across them. This effect is most prominent for the 'minor' contributor F8, in Mix4 **G03_F_Mix04_0.25_2to1_a.** The log(*LR*) drops from approximately 23 to 11 through the addition of an extra contributor. In this example the mixture proportions change from approximately 73:27 when interpreted as a two person mix to 45:34:20 when interpreted as a three person mixture. The impact on the weights for each genotype combination is subsequently quite significant. Due to the additional genotype combinations STRmix™ accepts, overestimating the number of contributors resulted in a shift of exclusionary *LRs* of 0 (plotted at -40) to non-zero *LRs* ranging between a log(*LR*) of approximately -35.00 to 3.0 for non-contributors. This is reflected in the group of H_d true values clustered on the far left of the *x*-axis.

Broadly speaking, over-estimating the number of contributors can result in false inclusions of non-contributors, however these are often very low *LR*s.

Subtraction of one contributor

The assumption of one fewer contributor than that actually present may be made when contributors are at very low levels and dropping out or there is allele sharing, perhaps due to relatedness.

During the analysis of the profiles in Section D, there were eleven four person mixtures that were run as three person (N-1) mixtures in STRMix™ (see Table 5).

Table 5: Samples originally analysed as experimental *N* contributors and then re-analysed as *N-1.*

Sample	Design NOC (*N*)	Apparent NOC (*N*-1)
B03_F_Mix08_1.0_10to5to2to1_b	4	3
B03_F_Mix08_1.0_10to5to2to1_a	4	3
G04_F_Mix08_0.25_10to5to2to1_a	4	3
B05_F_Mix08_0.1_9to3to3to1_a	4	3
F05_F_Mix08_0.1_9to3to3to1_a	4	3
F05_F_Mix08_0.1_9to3to3to1_b	4	3
G04_F_Mix08_0.25_10to5to2to1_b	4	3
G05_F_Mix08_0.1_6to3to1to1_a	4	3
C05_F_Mix08_0.25_4to3to2to1_b	4	3
G05_F_Mix08_0.1_6to3to1to1_b	4	3
B06_F_Mix08_0.1_1to1to1to1_a	4	3

Similar to above the *LR* for the *N* known contributors and the 500+*N* known non-contributors (as used for the specificity and sensitivity studies, Section D) were calculated. The log(*LR*) was compared for the known contributors and known non-contributors under the assumption of *N* and *N*-1 contributors. A plot of log(*LR*) of *N*-1 and experimental design *N* is provided in Figure 6.

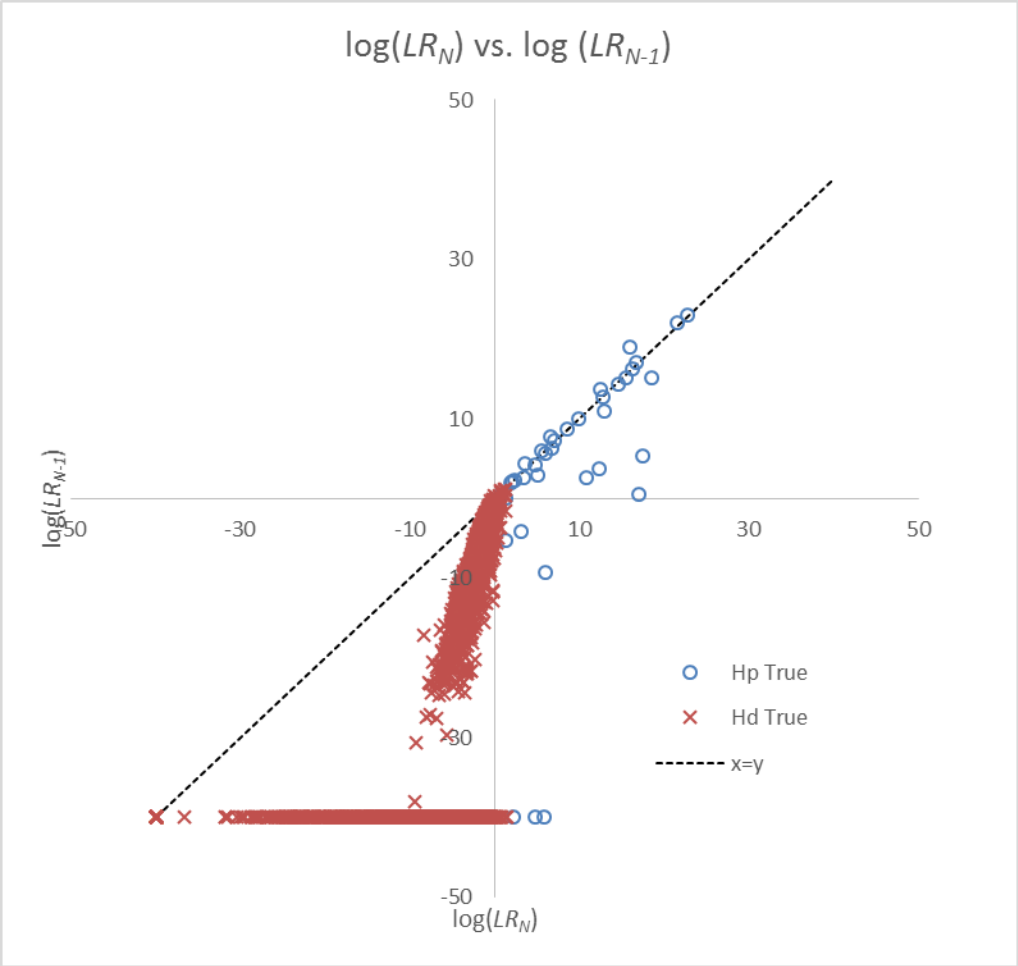


Figure 6: log(*LR*) when *N* is under-assigned (*N*-1) versus log(*LR*) for experimental design *N* for the known and non-contributors (where H_p true are in blue circles and H_d true in red crosses).

The plot above shows that for the majority of the H_p true *LR*s there was little effect. However, it also demonstrates that when the number of contributors is under estimated false exclusions of true contributors can be observed. This is because by under-assigning the number of contributors to a mixture, STRmix™ is restricted in its ability to propose and accept genotype combinations and these may not align to the true minor or trace contributor. Table 6 shows the impact on five individuals who were excluded when run as three contributors (*N*-1) compared to log(*LR*) obtained when the mixtures were run at the experimental design of four or *N* contributors. Note two of these examples also had *LR*'s of less than one (supporting H_d under the experimental design of four contributors.

Table 6: **Summary of the log(*LR*) values for previously excluded H_p true individuals in Section D.**

Sample	donor	AppNOC (*N*-1)	Log(*LR*)	Exp NOC (*N*)	Log(*LR*)
C05_F_Mix08_0.25_4to3to2to1_b	F19	3	-40	4	5.814
G04_F_Mix08_0.25_10to5to2to1_b	F19	3	-40	4	4.761
B06_F_Mix08_0.1_1to1to1to1_a	M36	3	-40	4	2.182
B03_F_Mix08_1.0_10to5to2to1_b	F19	3	-40	4	-0.361
F05_F_Mix08_0.1_9to3to3to1_b	F24	3	-40	4	-1.861

Under-estimating the number of contributors results in lower *LR*s for H_d true comparisons as STRmix™ is not having to explain any additional trace components to the mixture as potentially allelic. It should be noted that STRmix™ presents the user with a warning message and will not run if there are peaks present that cannot be explained using stutter modelling, or drop-in and can only be explained via an extra contributor being present in the mixture.

Section G: Allele drop-in

This section covers the following standard:

4.1.8 Allele drop-in

The LR should not change when drop-in peaks are within model paramters. For alleles that are outside of model parameters, a LR of zero should be returned or the interpretation in STRmix™ should not progress as the profile can no longer be explained by the assumed number of contributors. This was demonstrated in the previous validation of V2.4.

Section H: Stutter

This section covers the following standard:

4.1.9 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, the allele designation itself, or the average observed stutter ratio for the allele. STRmix™ can also implement either a per allele or per locus stutter model for any stutter type. Stutter peak are retained at analysis and within the STRmix™ input file. Stutter peaks (those types that are being modelled) should be retained at analysis and exported to the STRmix™ evidence 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.

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 $c^2 = 7.884$ (the 75th percentile from the allelic variance prior distribution for this data set). The 95% bounds encapsulate sufficient data as demonstrated in the graphs (coverage = 96.7%) demonstrating that the values for variance are sufficiently optimised (Figure 7).

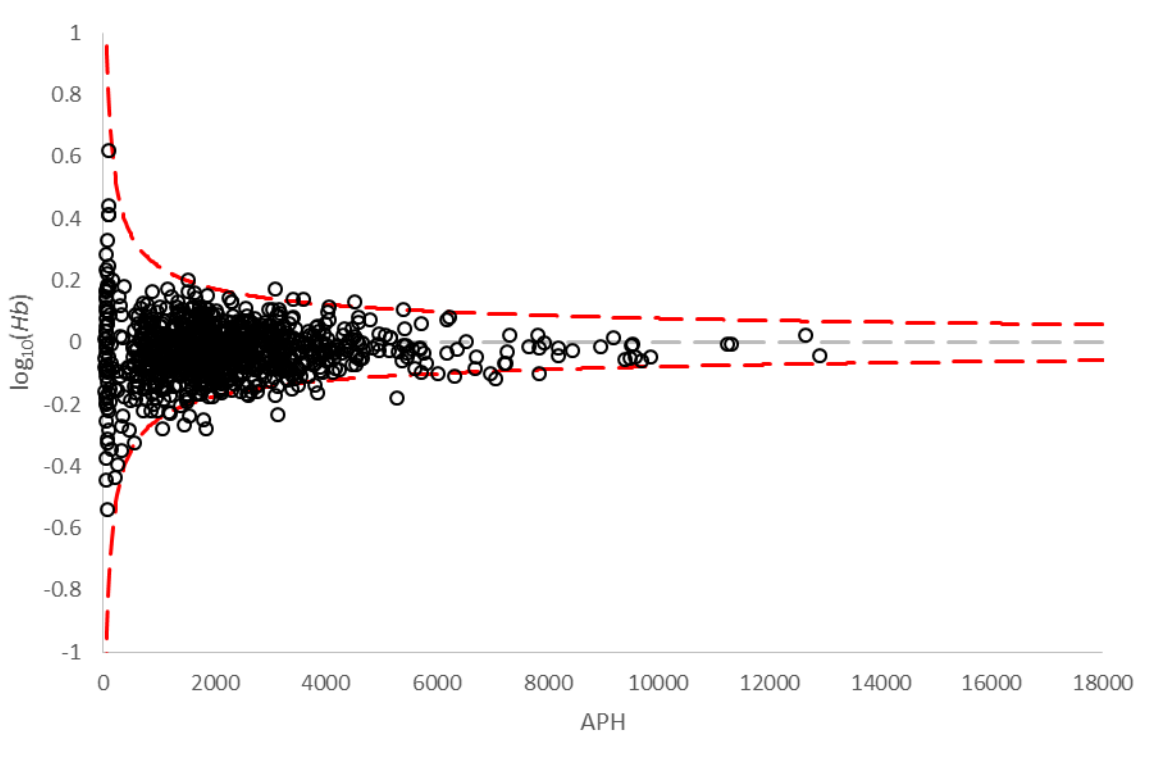


Figure 7: Log(*Hb*) versus APH for the single source profiles used in Model Maker at the PBSO laboratory.

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.

This principle was demonstrated in the previous validation of V2.4.

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.

This principle was demonstrated in the previous validation of V2.4.

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 formula.

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

This principle was demonstrated in the previous validation of V2.4. In addition, PowerPlex Fusion 6C is a new STR typing kit that will be implemented in the lab. An interpretation model other than a continuous probabilistic approach is not available for comparison.

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 Markov Chain Monte Carlo (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 laboratory variables. The variability due to the size of the allele frequency database and the MCMC process is taken into account within STRmix™ 2.6 using the highest posterior density (HPD) method [10-12] (a type of confidence interval).

This principle was demonstrated in the previous validation of V2.4 and demonstrated by the sensitivity and specificity studies covered above in Section D.

The use of replicate amplifications

STRmix™ allows for the interpretation of multiple PCR replicates in one interpretation, even if different amounts of template DNA have been added to the PCRs. The model assumes that the replicates originate from the same DNA extract and are not, for example, repeat injections of amplified DNA. The use of replicates has been shown to improve the ability of STRmix™ to differentiate true from non-contributors, generally increasing the *LR*s for true contributors and decreasing the *LR* for non-contributors (8).

There is a convincing argument against splitting the DNA extract to allow for multiple replicates[1]. However, assuming sufficient extract remains after initial amplification, there have been numerous reports of two replicates increasing the information content with regard to a single DNA interpretation, thus providing more data for making inferences about the donor's genotype.

The purpose of this experiment was to determine the effect of replicate interpretation on subsequent *LR*s.

Replicate amplifications were undertaken as part of the validation plan (Section D) and were analysed individually. A selection of ten of the samples from the three and four person mixture series were re-analyzed in STRmix™ using both replicates within the same deconvolution. These samples are listed in Table 7.

Table 7. Samples used in the replicate study

Mixture	NOC
Mix8 9:3:3:1 A04 0.5ng (9-10)	4
Mix8 6:3:1:1 A04 0.25ng (13-14)	4
Mix8 4:3:2:1 A06 0.1ng (16-17)	4
Mix8 10:5:2;1 B03 1ng (34-35)	4
Mix8 4:4:1:1 B05 0.25ng (42-43)	4
Mix6 10:5:1 D09 - E04 1ng (101-115)	3
Mix6 1:1:1 C11 - F11 0.1ng (80-152)	3
Mix6 8:1:1 A10-D10 0.5ng (24-103)	3
Mix6 3:2:1 B10-E10 0.5ng (52-129)	3
Mix6 1:1:1 B11 - G10 0.25ng (54-175)	3

An *LR* was calculated for the known donors to these mixtures using the same allele frequencies, theta, and *LR* based on experimental N, using Database Search within STRmix™ given 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

The log(*LR*) calculated for replicate and single profile interpretations are displayed in the plot below (Figure 8). Each of the two replicates for each sample has been compared to the outcome of the analysis of both replicates. Values above the line at *x*=*y* indicate an increase in the *LR* when using replicates. For the most part, the use of replicate amplifications increased the LR ratio for true contributors while decreasing the LR for non-contributors.

[1] Todd Bille and Michael Coble, Comparison of the deconvolution and likelihood ratios produced using STRmix™ software from low level samples when amplifying the entire extract or splitting the extract. ANZFSS 23rd International Symposium on the Forensic Sciences, Auckland, New Zealand, September 2016

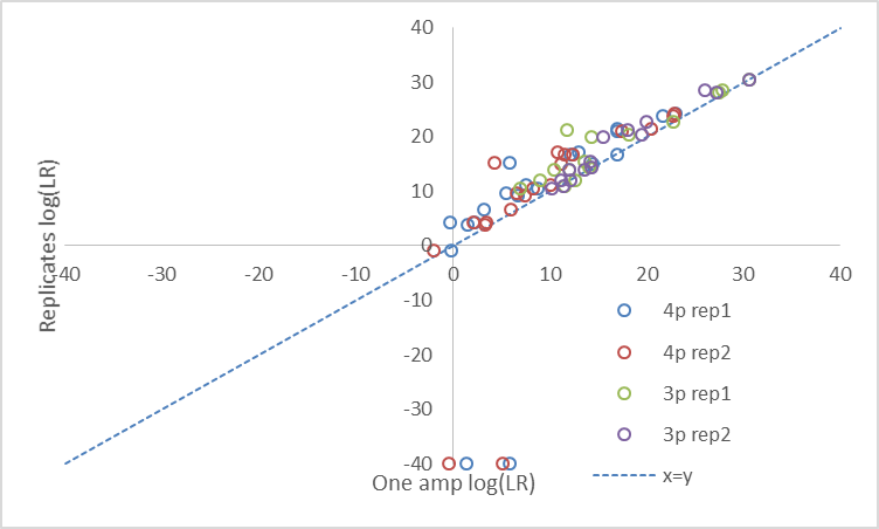


Figure 8. **The impact of replicate analyses on the log(LR) of true donors to mixtures**

One individual (F19) was excluded (plotted as log (LR) of -40), as a contributor in two of the four person mixtures when replicate amplifications were combined when previously they were not excluded from the four individual deconvolutions. The two four person mixtures were; **B03_F_Mix08_1.0_10to5to2to1** replicates **a** and **b,** and **A04_F_Mix08_0.5_9to3to3to1** replicates **a** and **b**. The individual excluded (F19) was the third contributor to each sample but not the lowest level contributor. When these mixtures were compared directly to the reference profile for F19 the exclusion was shown to be at locus D1S1656. F19 has the genotype 12, 15.3 at this locus. Inspection of the four e-grams for each of the individual replicates shows that in each case the 15.3 peak has not been fully resolved from the larger 16 peak but is clearly visible. (See Figure 9.)

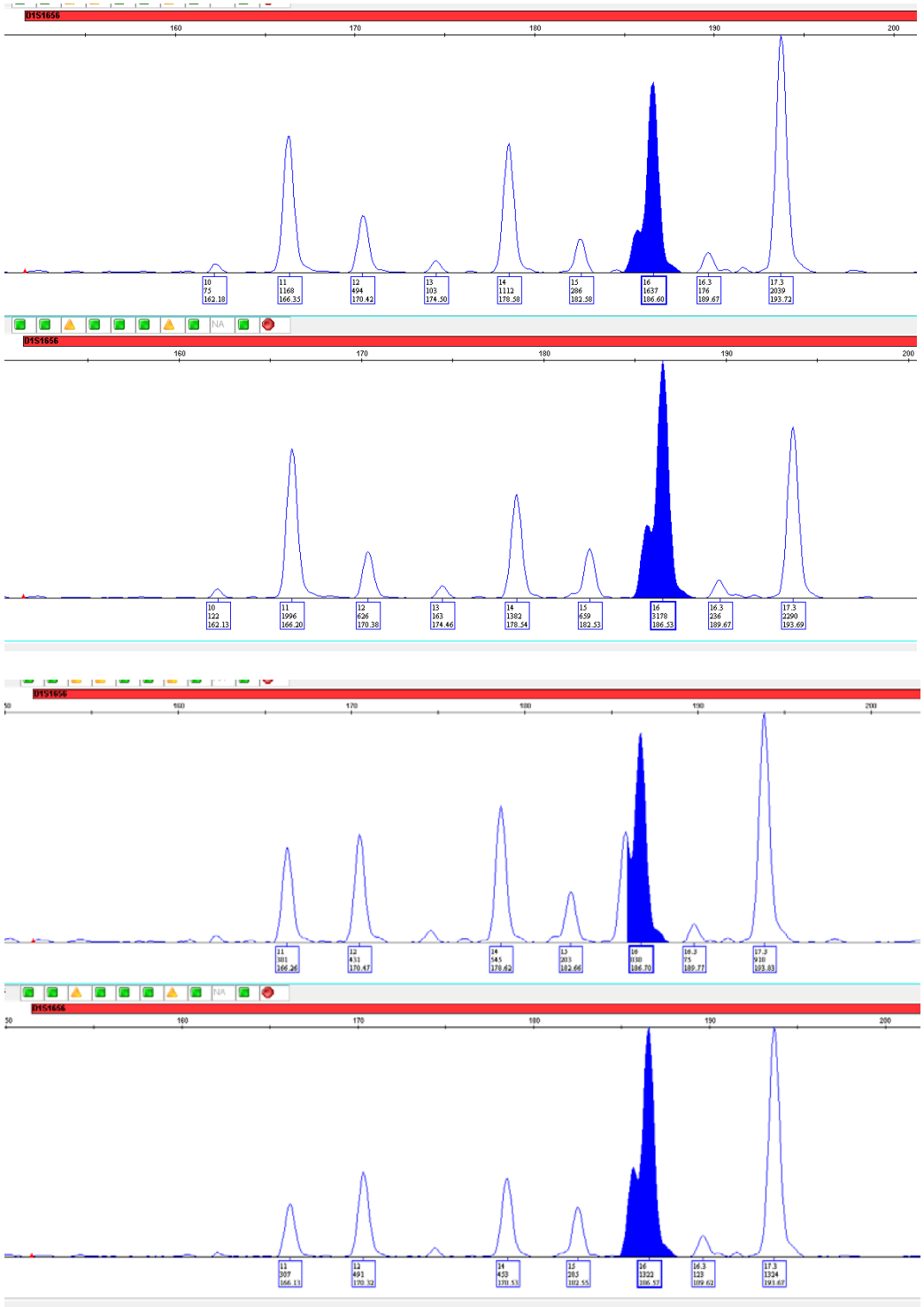


Figure 9. **E-gram images of locus D1S16156 for B03_F_Mix08_1.0_10to5to2to1 replicates a and b (upper two panes) and A04_F_Mix08_0.5_9to3to3to1 replicates a and b (lower two panes).**

When the LR's for F19 were recalculated from the replicate analyses ignoring D1S1656 strong inclusionary LR's were obtained for both mixtures. When the replicates were analysed individually, STRmix™ accepted genotype combinations that included drop-out meaning that F19 was not excluded. Combining replicates provided more data for the peak heights of all the other alleles detected. Given the heights of the other peaks attributed to this contributor two separate drop out events of 15.3 are considered much less likely and consequently this genotype combination has not been accepted during MCMC. This is not surprising given the presentation of the peaks at the D1S1656 locus in the examples above.

Validation of VarNOC feature of STRmix™ v2.6.2

The variable number of contributors (hereafter varNoC) function is a new feature in STRmix™ v2.6 onwards. Conventionally, when setting up an analysis in STRmix™, the user is required to input the apparent number of contributors (*N*) to the profile being interpreted. There may be occasions where *N* cannot be assigned with confidence; in these situations, STRmix™ allows for a profile to be interpreted using a range of values for *N*. However, it is recommended that the varNoC function is only used in casework after all other avenues to reduce the uncertainty in assigning *N* have been explored. While any range can be entered, developmental validation of the varNoC function has only been carried out for a contributor range of (+/-) one. If an increased range is required, it could be argued that too much uncertainty exists in the profile to progress a meaningful interpretation.

In this section the effect of varNoC interpretation on subsequent *LR*s was examined. Fourteen mixtures where there was some ambiguity in the NOC and a difference between the analyst assigned apparent NOC and experimentally designed NOC were re-interpreted using a contributor range *N*->*N*+1. Here *N* indicates the lower estimate for NOC. The mixtures selected, the experimental design and apparent number of contributors, and the range of contributors used in VarNoC calculations are displayed in Table 8.

Table 8: **Summary of mixtures interpreted using varNoC. Experimental design *N* was the value of NOC used in Section D. Apparent *N* indicates the interpretation made for NOC.**

Sample	Experimental design *N*	Apparent *N*	VarNoC Range
F05_F_Mix08_0.1_9to3to3to1_b.hid.csv	4	3	3;4
G04_F_Mix08_0.25_10to5to2to1_a.hid.csv	4	3	3;4
G04_F_Mix08_0.25_10to5to2to1_b.hid.csv	4	3	3;4
G05_F_Mix08_0.1_6to3to1to1_a.hid.csv	4	3	3;4
G05_F_Mix08_0.1_6to3to1to1_b.hid.csv	4	3	3;4
C03_F_Mix04_0.25_10to1_b.hid.csv	2	3	2;3
G03_F_Mix04_0.25_2to1_a.hid.csv	2	3	2;3
G04_F_Mix04_0.1_1to2_b.hid.csv	2	3	2;3
B03_F_Mix08_1.0_10to5to2to1_a.hid.csv	4	3	3;4
B03_F_Mix08_1.0_10to5to2to1_b.hid.csv	4	3	3;4
B05_F_Mix08_0.1_9to3to3to1_a.hid.csv	4	3	3;4
B06_F_Mix08_0.1_1to1to1to1_a.hid.csv	4	3	3;4
C05_F_Mix08_0.25_4to3to2to1_b.hid.csv	4	3	3;4
F05_F_Mix08_0.1_9to3to3to1_a.hid.csv	4	3	3;4

The database search function was used in the same manner as section D to calculate an *LR* for each individual in the database; both contributors and non-contributors, with an *LR* threshold of 0, however a varNoC stratified *LR* was calculated rather than a 'standard' *LR*. The proposition sets used in this section of the analysis were:

H_p: The DNA originated from the database individual and *the VarNoC range of N minus 1* unknown individuals
H_d: The DNA originated from the *VarNoC range of N* unknown individuals

The NIST Caucasian allele frequencies with a theta (F_{ST}) of 0.01 (1%) were used for Database Search *LR* calculations. The (standard, non-VarNoC) *LR*s calculated in Section D were compared with the varNoC *LR*s and are plotted in Figure 10.

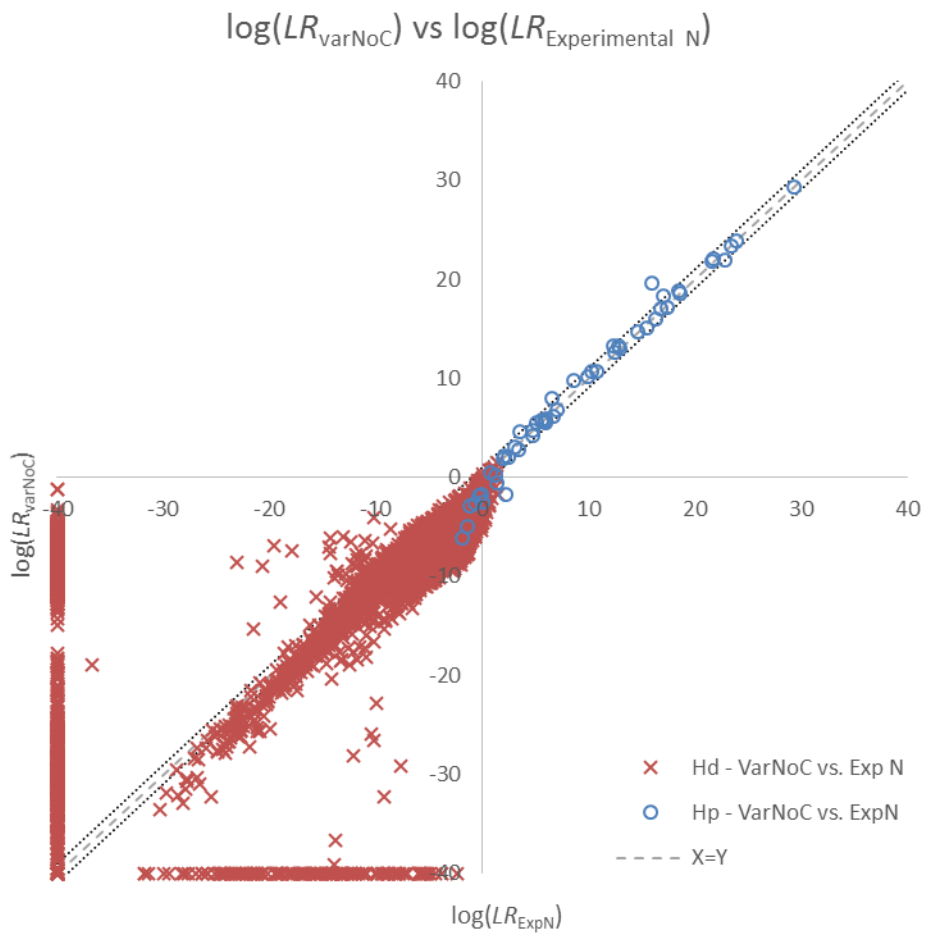


Figure 10: Comparison of the log(varNoC *LR*) and the non-VarNoC log(*LR*). True contributors are shown as blue circles and known non-contributors are shown as red crosses. The dashed line indicates the $x = y$ trend line, and dotted lines indicate a ±1 order of magnitude.

As shown in Figure 10, the results of the known contributors show good concordance between the varNoC *LR*s and the *LR*s calculated using a single value for the number of contributors' assignment. Figure 10 shows that 36 of the 50 observations (72%) for the true contributors exhibited *LR*s that were within one-order of magnitude of the *LR* obtained from VarNoC deconvolution methods. This is demonstrated by the majority of the blue circles falling in-between the black dotted lines that indicate ±1 order of magnitude difference from the $x = y$ line. Forty-five of the fifty samples (90%) were within two-orders of magnitude. Overall, these results demonstrate the robustness of the *LR*s calculated for the known contributors to each mixture when deconvoluted using a range of contributors. Any variability can be attributed to not only MCMC variation, but also the varNoC deconvolution process.

The varNoC interpretations show an increased variability in the *LR*s calculated for the non-contributors. This increasing variability is a result of spanning the range of contributors to apparent NOC to apparent NOC + 1. This is because STRmix™ adds an additional unseen profile at low DNA template amounts, diffusing some of the genotype weights allowing for more genotype combinations, with low genotype weights. This is a similar trend observed in Section F.

Conclusion

This document describes the PBSO's internal validation activities for STRmix™ V2.6. It has been shown that it is suited for its intended use for the interpretation of Fusion 6C® profiles generated from casework samples.

Signatures

This work has been reviewed and it has been determined that STRmix™ V2.6 is suitable for its intended use for interpretation of casework samples at PBSO. The project work has met the validation requirements as required by ANAB, The FBI Quality Assurance Standards, and the SWGDAM Validation Guidelines for the Validation of Probabilistic Genotyping Systems.

Amy McGuckian, FBU Technical Leader

References

(1) J.-A. Bright, D. Taylor, C. McGovern, S. Cooper, L. Russell, D. Abarno, J. Buckleton, Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles, Forensic Science International: Genetics 23 (2016) 226-239.

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

(3) J.-A. Bright, I.W. Evett, D. Taylor, J.M. Curran, J. Buckleton, A series of recommended tests when validating probabilistic DNA profile interpretation software, Forensic Science International: Genetics 14 (2015) 125-131.

(4) D.J. Balding, R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, Forensic Science International 64 (1994) 125-140.

(5) National Research Council II, National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence, National Academy Press, Washington, D.C., 1996.

(6) M. John M. Butler. Genetics and Genomics of Core Short Tandem Repeat Loci Used in Human Identity Testing J Forensic Sci, Vol. 51, No. 2.

(7) l. Sez. V, Massimario N°23, La Giustizia Penale 35(1 (Gennaio)) (1994) 42.

(8) D. Taylor, Using continuous DNA interpretation methods to revisit likelihood ratio behaviour, Forensic Science International: Genetics 11 (2014) 144-153.

(9) 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 Science International: Genetics 12 (2014) 208-214.

(10) J.-A. Bright, D. Taylor, J. Curran, J. Buckleton, Searching mixed DNA profiles directly against profile databases, Forensic Science International: Genetics 9 (2014) 102-110.

(11) J.-A. Bright, K.E. Stevenson, J.M. Curran, J.S. Buckleton, The variability in likelihood ratios due to different mechanisms, Forensic Science International: Genetics 14 (2015) 187-190.

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

(13) C.M. Triggs, J.M. Curran, The sensitivity of the Bayesian HPD method to the choice of prior, Science & Justice 46(3) (2006) 169-178.

(14) J.M. Curran, J.S. Buckleton, An investigation into the performance of methods for adjusting for sampling uncertainty in DNA likelihood ratio calculations, Forensic Science International: Genetics 5(5) (2011) 512-516.

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 behaviour. 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.
7. Taylor D, Buckleton J. Do low template DNA profiles have useful quantitative data? Forensic Science International: Genetics. 2015;16:13-6.
8. Taylor D, Buckleton J, Bright J-A. Factors affecting peak height variability for short tandem repeat data. Forensic Science International: Genetics. 2016;21:126-33.

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.
4. Moretti TR, Just RS, Kehl SC, Willis LE, Buckleton JS, Bright J-A, et al. Internal validation of STRmix™ for the interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics. 2017;29:126-44.
5. Bright J-A, Taylor D, McGovern CE, Cooper S, Russell L, Abarno D, et al. Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles. Forensic Science International: Genetics. 2016;23:226-39.
6. Taylor D, Bright J-A, McGoven C, Hefford C, Kalafut T, Buckleton J. Validating multiplexes for use in conjunction with modern interpretation strategies. Forensic Science International: Genetics. 2016;20:6-19.
7. J.-A. Bright et al. Internal validation of STRmix™ – A multi laboratory response to PCAST. Forensic Science International: Genetics, 2018. 34: 11-24.

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

Appendix 3: Review of Diagnostics from Section D

This section summarizes the secondary diagnostics for the mixtures in Section D. These include the total number of iterations, log(likelihood), Gelman-Rubin convergence diagnostic, and the posterior mean allele and stutter variances. Secondary diagnostics are a useful guide to provide confidence the interpretation has progressed as expected. Individual secondary diagnostics may indicate whether a further scrutiny is warranted, however analysts should not rely on these diagnostic alone. Further review of the other diagnostics and the profile itself could indicate that STRmix™ is performing as expected.

Total Number of Iterations

The total number of iterations simply shows the number of iterations required for 400,000 accepts (50,000 accepts across eight MCMC chains) to be reached. This is the required default number of accepts for a DNA profile to complete deconvolution in STRmix™ V2.6. As shown in Figure 11, the number of iterations is expected to increase as DNA profiles become more complex. Excessive number of iterations could indicate that STRmix™ could not converge on good probability space during MCMC, speaking to the complexity of the profile.

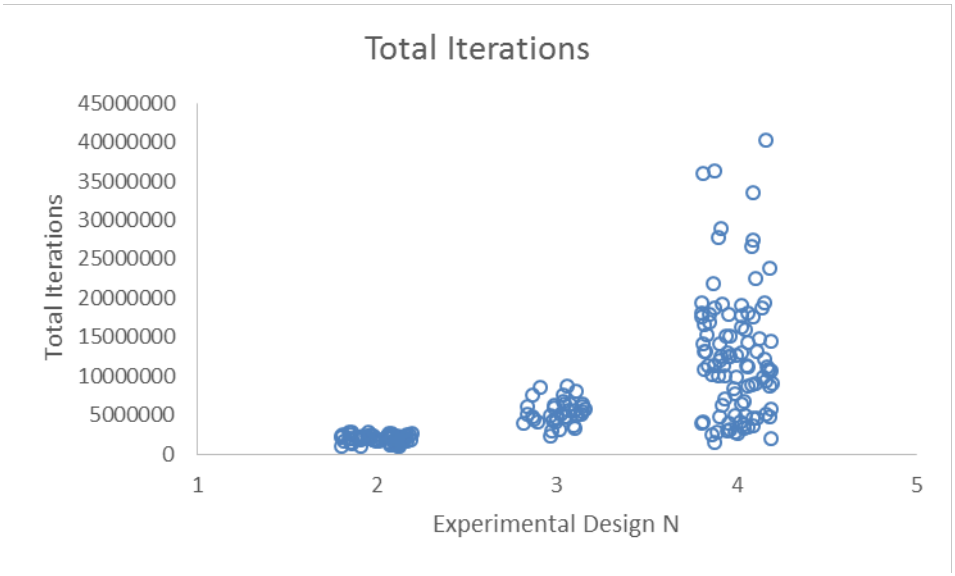


Figure 11: The total number of iterations required for each mixture to yield 400,000 accepts.

As expected as the complexity of the profiles increases a general increase in the total number of iterations is observed.

Average log(likelihood)

The average log(likelihood) can be described as the average post burn-in probability density (or likelihood) values across the chains used in a deconvolution. The values shown in Figure 12, (with the exclusion of one

sample discussed below) show a spread of these values, ranging from close to 0 to >80. Generally, a high average log(likelihood) is better as it indicates that STRmix™ has been better able to model the expected profile compared to the observed. However, low or negative values, do not necessarily indicate an issue. These may be observed in very partial profiles where there is very little peak height information contributing to the overall calculation of the log(likelihood).

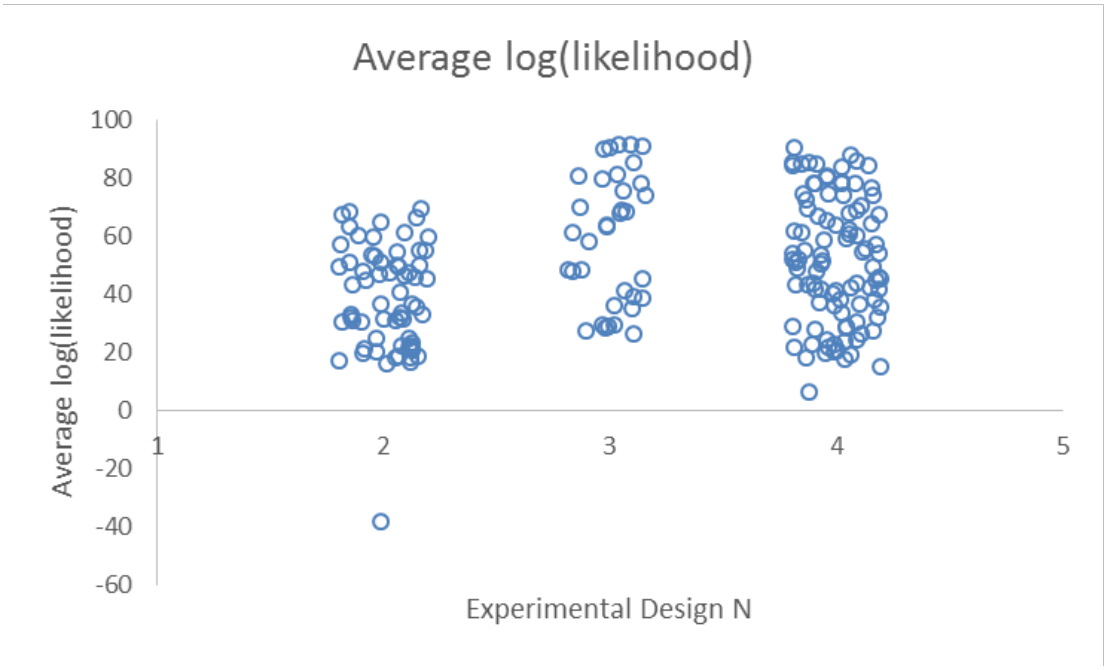


Figure 12: The average log(likelihood) output for each mixture.

For example, there is one notable outlier within the two person mixture set (Mix 4), sample **C01_F_Mix04_1.0_5to1_a** where a negative log(likelihood) was observed. A review of this input file and DNA profile revealed the absence of a stutter peak, 10.3 at D2S441 from a large 11.3 allele. This stutter peak appears as an unresolved shoulder to an 11 allele peak from the second contributor. This is also displayed at D12S391, where 16.3 stutter peak is not labelled from a 17.3 allele, where a 17 allele is also present. Figure 13 displays excerpts from the electropherogram of this sample. Depending on the expected profile proposed during MCMC, the absence of a stutter peak or peaks in the observed profile can lead to a 'penalty' being applied during the calculation of the log(likelihood). The mixture proportions, weights and *LR* values calculated for this mixture were all intuitive (meet quality control expectations). When a 10.3 peak is added 'in silico' to the input file for this mixture an intuitive log(likelihood) is obtained.

Other unintuitive secondary diagnostics were also observed in the Mix 4 mixture set due to the combination of alleles and corresponding stutter peaks at D2S441 and D12S391. These are discussed further below.

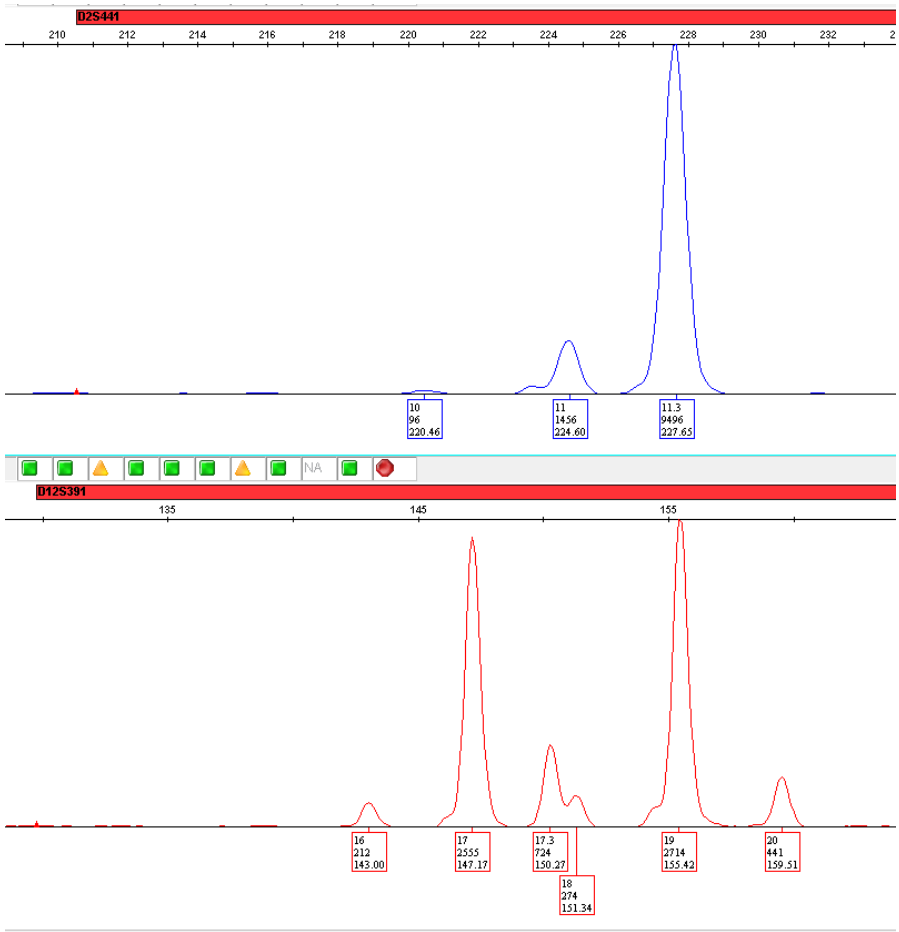


Figure 13: **E-gram at D2S441 (upper pane) and D12S391 (lower pane) for sample C01_F_Mix04_1.0_5to1_a**

Gelman-Rubin (GR) convergence diagnostic

The Gelman-Rubin (GR) is a diagnostic value that indicates whether there is convergence of the MCMC probabilities in each MCMC chain. This value is an average within and across all chains. Figure 14 shows the spread of GR values for the mixtures studied in Section D with the majority of the points below 1.2 (shown as the dashed line). A value of 1.2 or less typically indicates likely convergence of the MCMC chains.

However, GR values greater than 1.2 do not necessarily indicate that the deconvolution is unsuitable. This value could simply be indicating the complexity of the given mixture. When an excessive GR value is observed it is best to investigate the other primary and secondary diagnostics. If these other diagnostics are not within expectations, then the analyst may choose to re-run the sample with the same or an extended number of accepts. If this does not decrease the GR value, it may indicate a further review of the input file or re-assessment of N could be warranted.

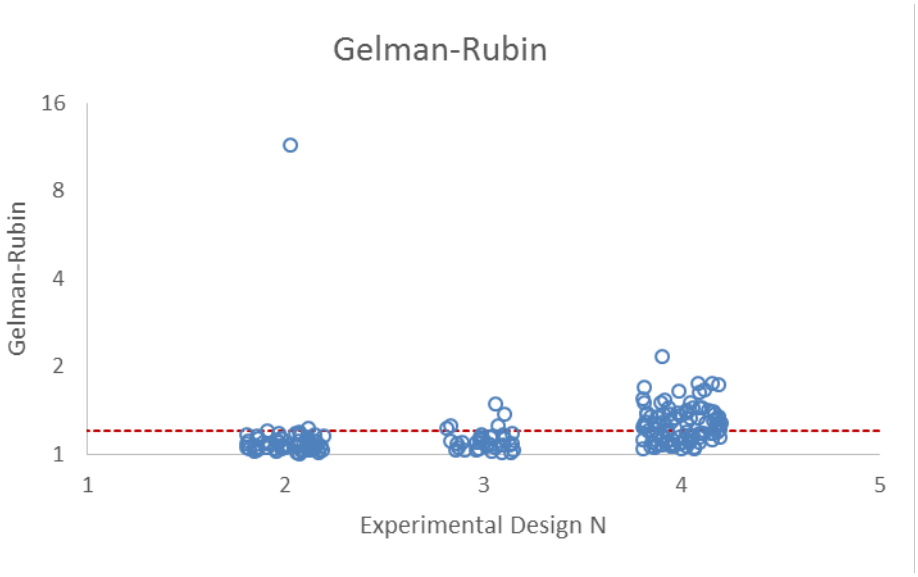


Figure 14: Summary of the GR values obtained for each mixture in Section D.

The one clear outlier, within the two person data set, was obtained from the 'b' replicate of the Mix 4 sample discussed above, **C01_F_Mix04_1.0_5to1_b**. A similar pattern of masked or missing stutters was observed at D2S441 and D12S391. Further investigation demonstrated that during the burn-in phase of this deconvolution some of the eight chains have followed a path where a 'penalty' for the absence of the 10.3 stutter at D2S441 has been applied and some have not. This deconvolution also exhibits high back stutter variance (see below) as would be expected with missing stutter peak labels as STRmix™ is having to propose a large variance value to accommodate the large difference between O (observed stutter) and E (expected stutter). The overall mixture proportions, weights and LRs are intuitive as they were for replicate 'a'. Addition of a 10.3 stutter peak into the input file at D2S441 gives rise to a GR value less than 1.2 and a reduced stutter variance. Deconvoluting this mixture with the original input file but using ten times more (1,000,000) burn-in accepts also results in a GR less than 1.2, but still with high stutter variance.

Posterior Mean Allele Variance

Figure 15 shows the spread of the allele variance values from the STRmix™ outputs for each sample run from Section D. This is calculated from the average of the accepted allele variance values from the 400,000 post burn-in accepts. The red, green and purple dotted lines represent the 50th, 75th and 90th percentile respectively of the allele variance prior distribution. The prior distribution plot is also provided as a point of reference. It can be seen that the majority of the posterior mean allele variances for each sample are situated around the main body of the allele variance prior distribution.

Occasionally a STRmix™ deconvolution of a DNA profile will lead to an inflated posterior mean allele variance, in that for whatever reason the profile requires a high allele variance.

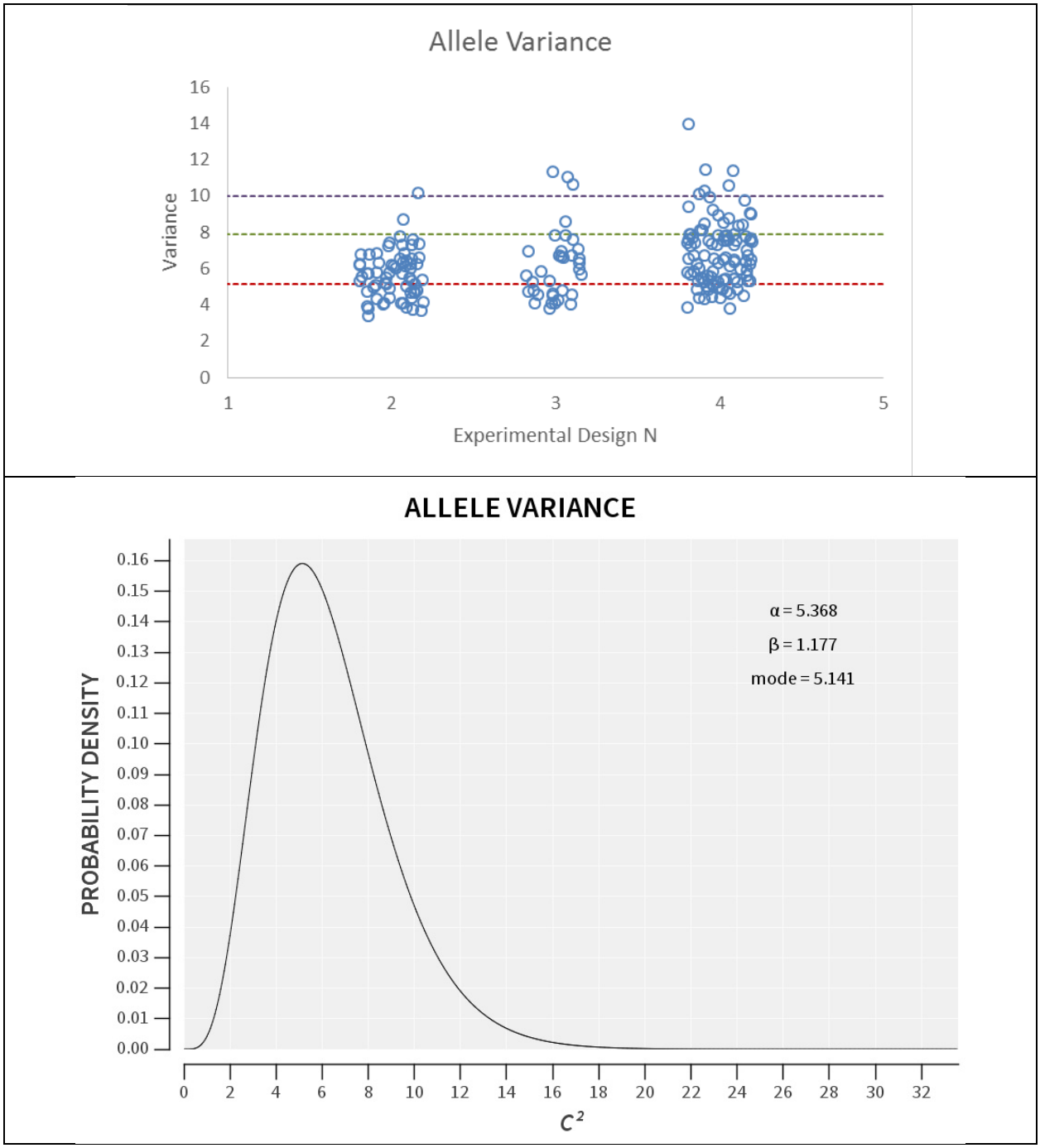


Figure 15: Plot of the posterior mean allele variance values for each DNA profile analyzed in Section D (upper pane) and the prior distribution modelled by a gamma distribution (lower pane).

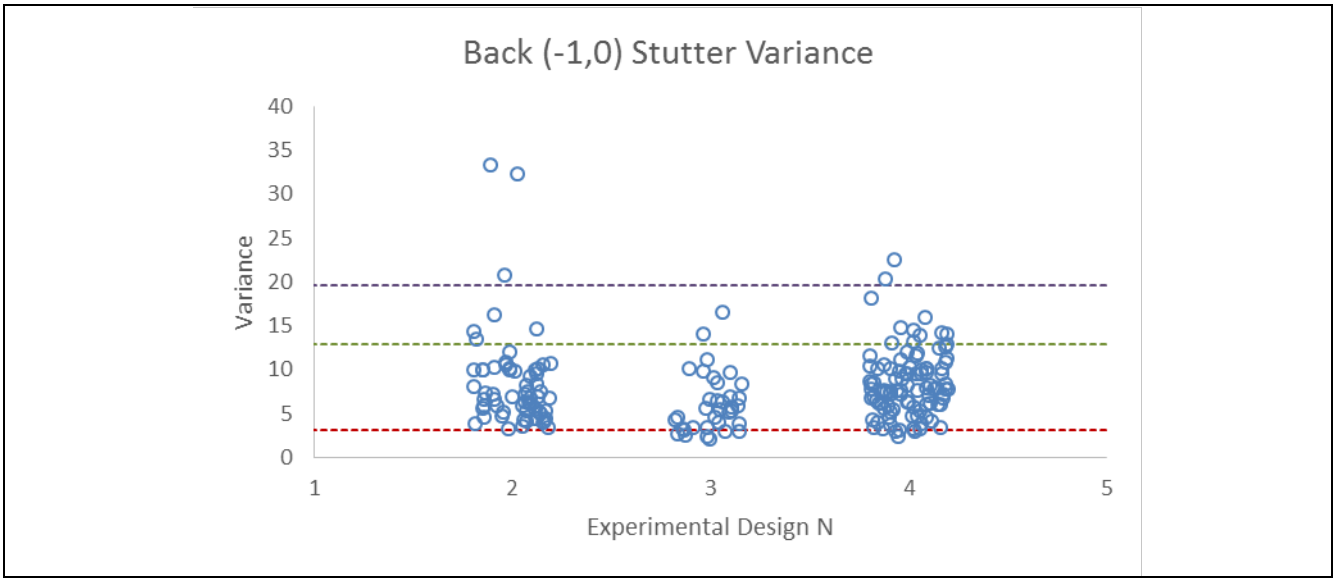
Posterior Mean Stutter Variance

Back Stutter (-1,0)

Figure 16 shows the spread of the back stutter variance values from the STRmix™ outputs for each sample run from Section D. This is calculated from the average of the accepted back stutter variance values from the 400,000 post burn-in accepts. The red, green and purple dotted lines represent the 50th, 75th and 90th percentile respectively of the back stutter variance prior distribution. Again the prior distribution plot is also provided as a point of reference. As seen in Figure 16, the majority of the posterior mean back stutter variances for each sample appears to be situated around the main body of the stutter variance prior distribution.

Occasionally a STRmix™ deconvolution of a DNA profile will lead to an inflated posterior mean back stutter variance, which is typically driven by the absence of stutters or stutters being larger or smaller than expected for various reasons. Two examples of elevated back stutter variance can be seen in the two person mixture set in Figure 16. One also had a high GR and was discussed above (**C01_F_Mix04_1.0_5to1_b**). The other (**E01_F_Mix04_1.0_1to2_a**) has the same combination of alleles and stutters but in this instance the absence of the 16.3 stutter at D12S391 is having the greatest impact on stutter variance.

An inflated posterior mean stutter variance does not indicate that the DNA profile is unsuitable for interpretation. It would simply indicate that the posterior mean stutter variance for this profile is outside the typical expectations of the prior distribution. Other diagnostic values and the profile itself should be considered to determine the suitability of this STRmix™ interpretation. As discussed above the primary diagnostics of mixture proportions, weights and *LRs* were all intuitive.



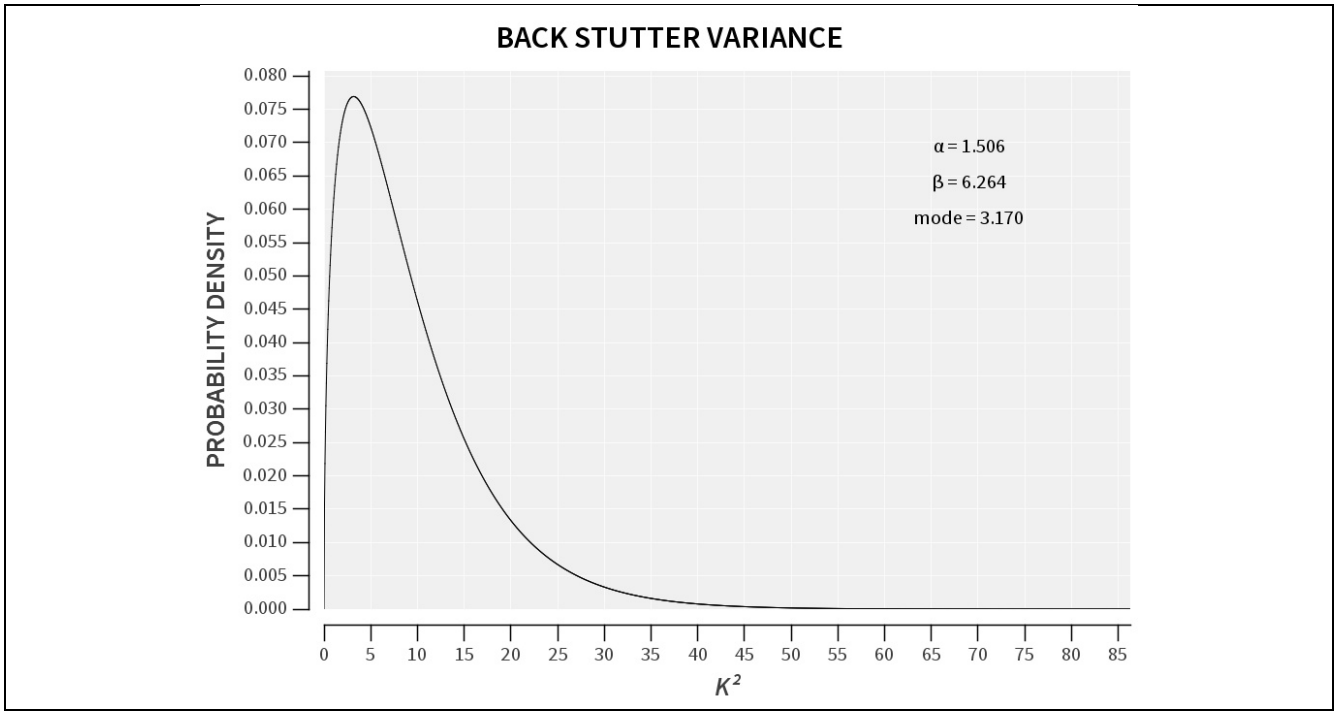
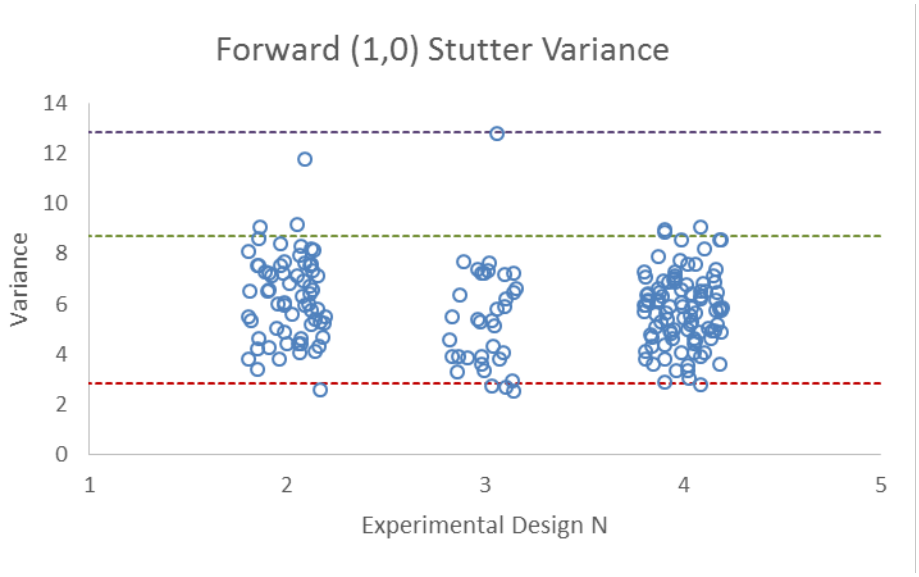


Figure 16: Plot of the posterior mean stutter variance values for each DNA profile analyzed in Section D (upper pane) and the prior distribution modelled by a gamma distribution (lower pane).

Forward Stutter (1,0), minus 2 base pair Stutter (-1,2) and Double Back Stutter (2,0) Variance

Figures 17, 18 and 19 show the spread of the forward stutter, minus two base pair stutter and double back stutter variance values, respectively, from the STRmix™ outputs for each sample run from Section D. These are calculated from the average of the accepted stutter variance values from the 400,000 post burn-in accepts. Red, green and purple dotted lines represent the 50th, 75th and 90th percentile respectively of the allele variance prior distribution. Again the prior distribution plots are also provided as a point of reference. As seen in Figures 17, 18 and 19, the majority of the posterior mean stutter variances for all three stutter types and each sample appear to be situated around the main body of the stutter variance prior distributions.



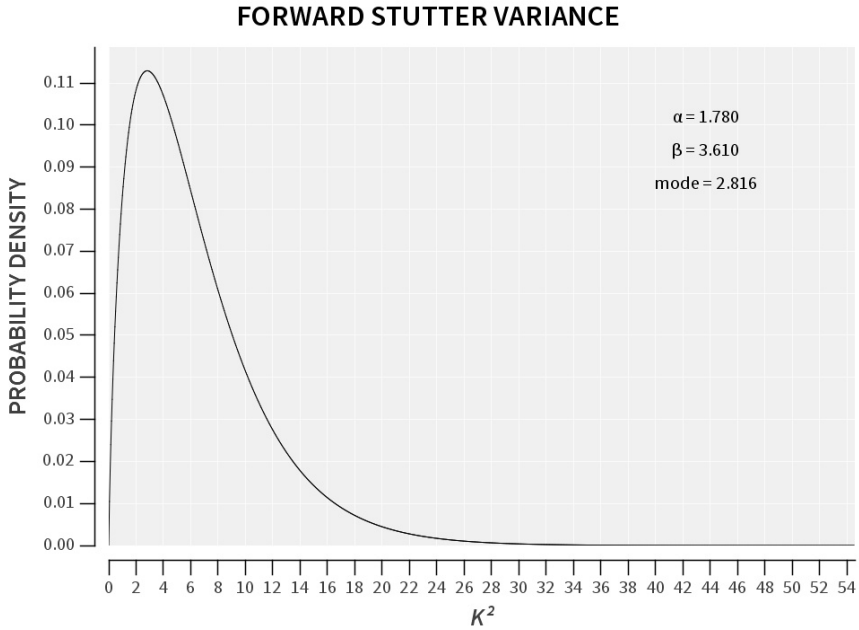
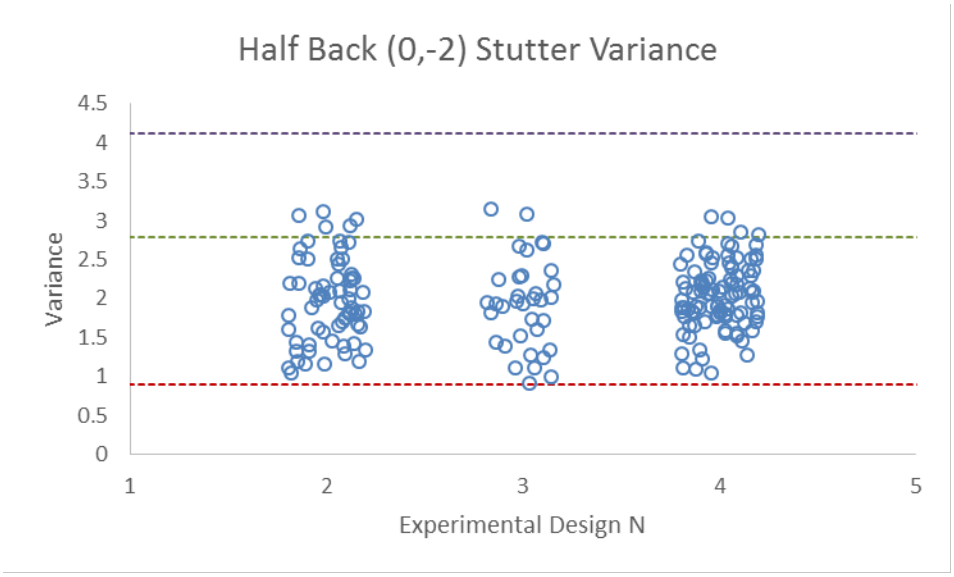


Figure 17: Plot of the posterior mean forward stutter variance values for each DNA profile analyzed in Section D (upper pane) and the prior distribution modelled by a gamma distribution (lower pane).



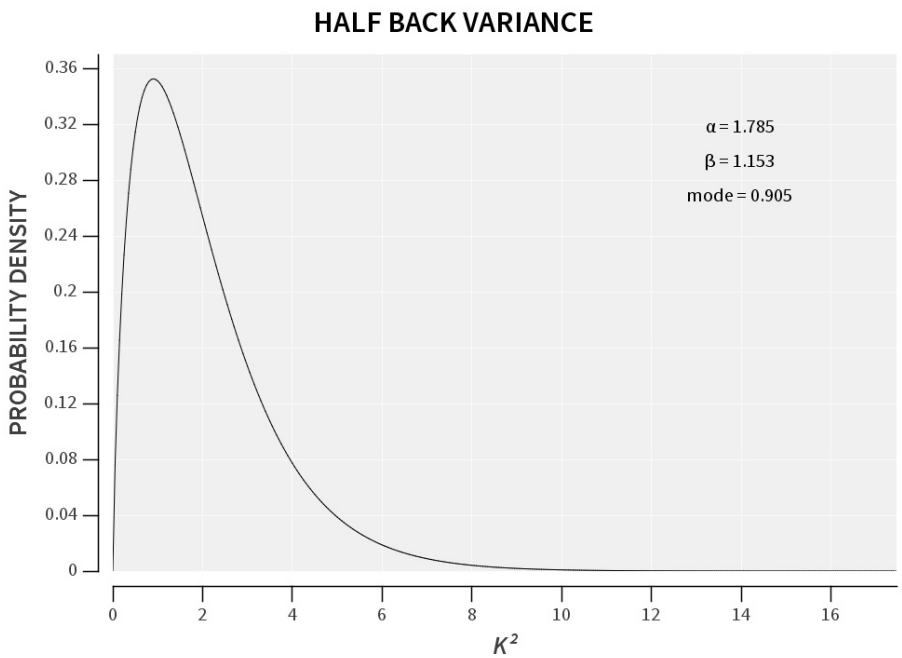
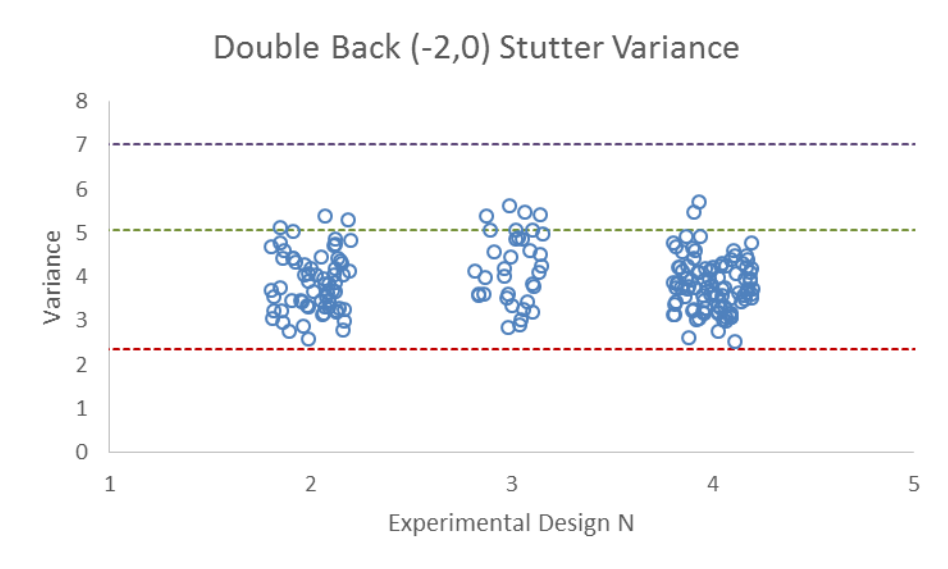


Figure 18: Plot of the posterior mean minus 2 base pair stutter variance values for each DNA profile analyzed in Section D (upper pane) and the prior distribution modelled by a gamma distribution (lower pane).



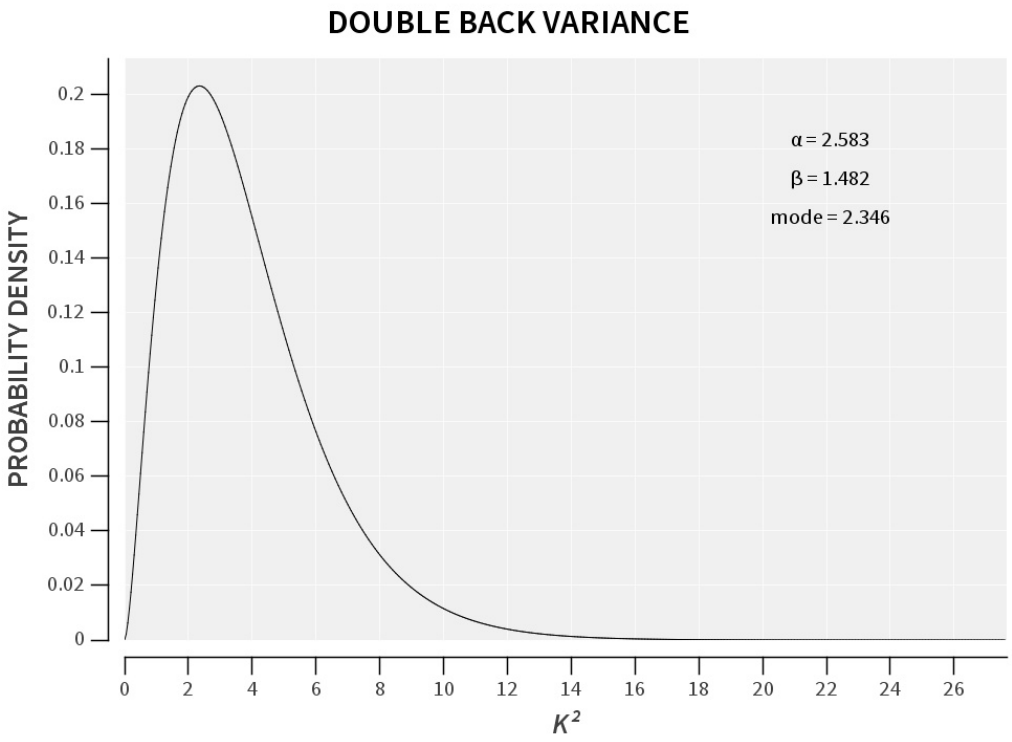


Figure 19: Plot of the posterior mean double back stutter variance values for each DNA profile analyzed in Section D (upper pane) and the prior distribution modelled by a gamma distribution (lower pane).