STRmix™ V2.9.1

(upgrade)

Idaho State Patrol Laboratory

(PowerPlex® 16, 3130 CE)

Introduction

This document describes the determination of laboratory-specific parameters within STRmix™ v2.9.1 for the Idaho State Patrol crime laboratory (hereafter, ISP). Parameters have previously been determined for STRmix™ v2.3 and PowerPlex®16 (PP16) data generated within ISP. For full detail of the original implementation and validation of STRmix™ for this kit configuration please refer to the implementation and validation reports held within the laboratory's quality management system. The data presented in this report is for an upgrade to STRmix™ V2.9.1 for PP16 data separated on Genetic Analyzer 3130 CE instruments.

Samples used to determine the laboratory-specific parameters and described in this report were all generated by ISP. Data analysis and work-up has been undertaken with the assistance of the STRmix™ support team from ESR in New Zealand.

STRmix™ parameters

There are a number of parameters which are not optimized by the MCMC in a STRmix™ analysis. These parameters must be set by the user and are either determined by analysis of empirical data or modelled within STRmix™ using Model Maker. The laboratory specific parameters that are determined prior to use of STRmix™ are:

- Analytical threshold (detection threshold)
- Stutter ratios
- Drop-in parameters
- Saturation
- Allelic and stutter peak height variance
- The hyper-parameter for the variance of locus specific amplification effects (LSAE).

These parameters need to be defined for each STR kit, each protocol (e.g. cycle number variation), and CE platform (e.g. 3130 or 3500), and potentially each time there is a significant change of platform (e.g. a camera or laser change). The analytical threshold, stutter settings (including stutter ratios), saturation settings and drop-in parameters[1] have been determined previously for ISP's PP16 data analyzed on 3130 capillary electrophoresis instruments using empirical data. These settings will be retained within the ISP PP16 STRmix™ V2.9.1. The peak height variances and locus specific amplification efficiencies for the PP16 dataset (3 different injection protocols) were re-evaluated using Model Maker within STRmix™ V2.9.1 from the analysis of empirical profile data. The results of these analyses are described within this report.

[1] Drop-in events have not been recorded for PP16 3130 data and therefore drop-in is not modelled within the STRmix™ PP16 3130 kit.

Peak height variance and LSAE using Model Maker

Empirical observations and experience suggests that profiles differ in variance (hereafter "quality"). Within STRmix™ the variability of peaks within profiles is described using a model containing a variance constant. Allele and stutter peaks have separate variances; c^2 and k^2, respectively. Furthermore, each stutter variant being modelled has its own k^2 variance constant. The c^2 and k^2 terms are variables which are determined after sampling from a gamma distribution within the MCMC.

The gamma distribution priors that STRmix™ samples from during an interpretation are optimised in Model Maker, an add-on to the STRmix™ software. Model Maker works by using a component wise MCMC. In component 1 each DNA profile has its mass parameters optimised and uses a stable gamma distribution for allele, stutter and LSAE variance constants. In component 2 the mass parameters for each profile are held constant and the hyperparameters for each gamma distribution are varied. Components are 1000 accepts long and they cycle through a number of times depending on the user input value.

To investigate the variance properties of the three different injection protocols (3, 5, and 10 second injection), 90 samples were prepared by ISP. The dataset is expected to be indicative of the peak height variability likely to be encountered in casework DNA profiles and was prepared by selecting a series of known donors and their DNA extracts diluted to create a range of input templates from 0.075ng to 2.5ng. These were amplified and run on the 3130 capillary electrophoresis instrument using each of the three injection protocols used within standard casework procedures. Each data set was run through Model Maker for 200 cycles (200,000 accepts total).

The resulting CE data was analyzed in GeneMapper™ using an analytical threshold of 50rfu across all dyes. Labels were retained for all allelic peaks and back stutter peaks. Following analysis, the Model Maker functionality of STRmix™ was used to assess peak height variability within the dataset. Whilst 90 samples were prepared not all samples were used in Model Maker. Profiles that contained 10 or fewer autosomal peaks or peaks in excess of the saturation threshold were not used. The number of samples analyzed and numerical representations of the prior distributions for each Model Maker run is displayed in Table 1. The variance parameters for each injection protocol are compared by overlaying the distributions for each variance type in Figure 1

Table 1: Details of the Model Maker analyses undertaken using data injected for 3, 5 and 10 seconds and summary of the details of the combined dataset comprising a subset of data from each injection protocol (full details on samples used is recorded within the combined dataset Model Maker run folder saved within the ISP records.

CE injection protocol	Number of samples analyzed	Allele variance parameters: α, β (mode)	Back stutter variance parameters: α, β (mode)	LSAE variance mean
3 seconds	82	3.122, 0.892 (1.893)	3.195, 1.271 (2.790)	0.008
5 seconds	85	4.021, 0.876 (2.646)	3.866, 1.585 (4.543)	0.008
10 seconds	79	6.528, 0.722 (3.991)	3.451, 2.125 (5.208)	0.009
Combined	80	3.949, 0.886 (2.613)	4.02, 1.432 (4.325)	0.009

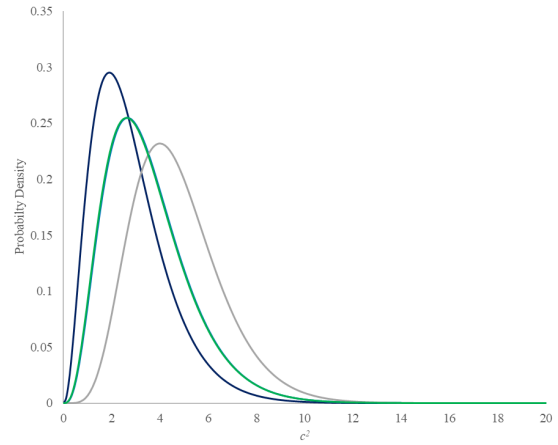
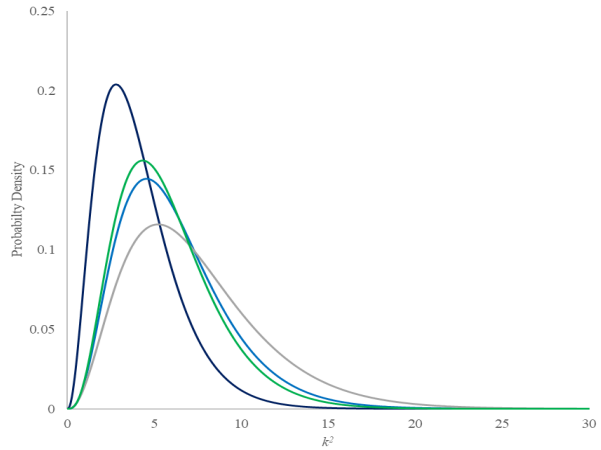
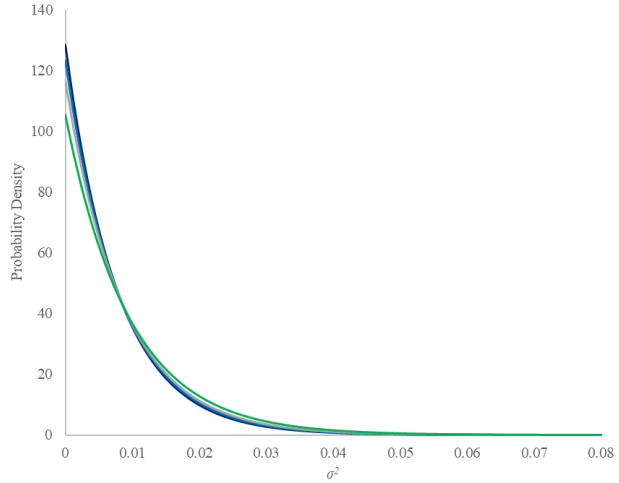
Allele variance	Back stutter variance
	
LSAE variance	
	

Figure 1: Peak height variance distributions for allelic and back stutter peaks, and the LSAE variance distribution for data developed using the 3, 5, and 10 second injection protocol plus a dataset comprising a subset of data from each. The dark blue distribution resulted from the Model Maker analysis of the 3 second data set, light blue the 5 second dataset, grey the 10 second dataset and green the combined dataset. Note: the 5 second distribution is obscured by the combined in the allele variance plot.

The modes and the coverage of the distributions are reasonably similar and sufficiently represented by the distributions of the combined dataset. In keeping with the STRmix™ V2.3 approach, the variance parameters derived from the combined dataset will be used within the ISP PP16 STRmix™ V2.9.1 kit.

The diagnostics output of the Model Maker runs were also reviewed. In particular, correlation plots were examined and assessed. The correlation plots of the combined dataset are presented as Figure 2. No obvious correlation was observed (the desired result). However, one outlier (present in the top left quadrant of the back stutter correlation plot) was investigated and found to be due to a lower-than-expected stutter of a vWA 14 allele in sample SH_1.5ng. Given the 14 allele peak height (3226 rfu) and the expected stutter ratio for a 14 allele (0.039892) in the ISP stutter exceptions file, STRmix™ expects a 13 stutter peak with height 128 rfu however, the observed height is only 59 rfu. This type of behavior at the vWA locus and specifically with 14 and 15 alleles has been observed in data from other laboratories. It has been observed there are at least two different 'populations' of vWA 14 and 15 alleles that exhibit different stutter ratios. This is likely due to different sequence variants that stutter in different amounts due to different underlying repeat structure[2]. STRmix™ is able to account for some degree of difference in the expected and observed peak heights however, scenarios such as described above with VWA 14 and 15 alleles may lead to potential over- or under-estimation of peak heights. Analysts need to be aware of this when reviewing the STRmix™ interpretation of profiles containing these alleles or with other allelic point variants that may not have been fully modelled in the stutter modelling phase.

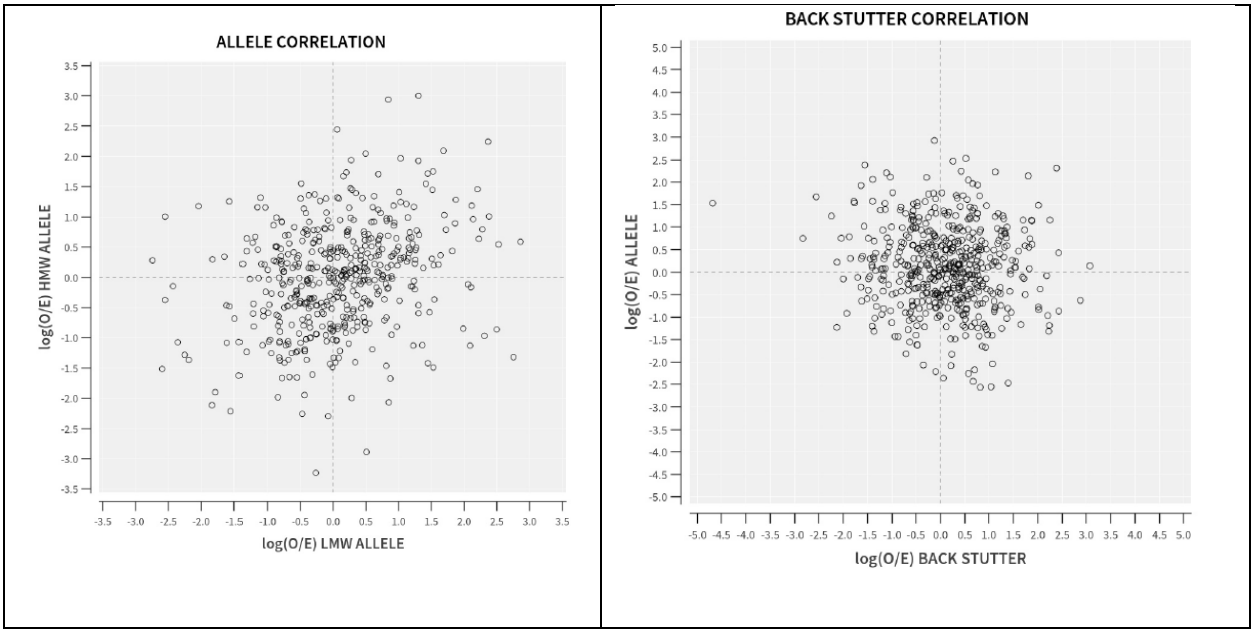


Figure 2*:* Correlation plots from the Model Maker analysis of the combined dataset

As a final check of the variance parameters determined, heterozygote balance was calculated for all heterozygote loci within the combined Model Maker dataset.

[2] This is a known issue and is further explained on the STRmix™ support website.

Heterozygote balance (*Hb*) was calculated as:

$$Hb = \frac{O_{HMW}}{O_{LMW}}$$

Where O_{HMW} refers to the observed height of the high molecular weight allele and O_{LMW} the observed height of the low molecular weight allele. Previous work has suggested that there is a relationship between the variation in peak height and the variation in *Hb* [1, 2]. In single source profiles, variability in *Hb* reduces as the average peak height (APH) at a locus increases. The variance of *Hb* can be used as a proxy for the variance of individual peaks. This allows an approximate comparison between the variance from the STRmix™ MCMC approach and a readily determined variable from empirical data (*Hb)*.

Plots of log(*Hb)* versus APH (the black circles) for the combined Model Maker dataset is provided in below. The expected 95% bounds are indicated within the plot using red dashed lines. The bounds were calculated as $\pm\sqrt{2}\times 1.96\times\sqrt{\frac{c^2}{APH}}$ where c^2 = 3.208 is the 50th percentile from the combined dataset allele peak height variance prior gamma distribution. Under the assumption of a normal distribution, it is expected that ~95% of data points fall within +/- 2 standard deviations (95% bounds) of the mean. For each dataset, the 95% bounds encapsulate sufficient data coverage = 96.2% demonstrating that the values for variance appear sufficiently optimized.

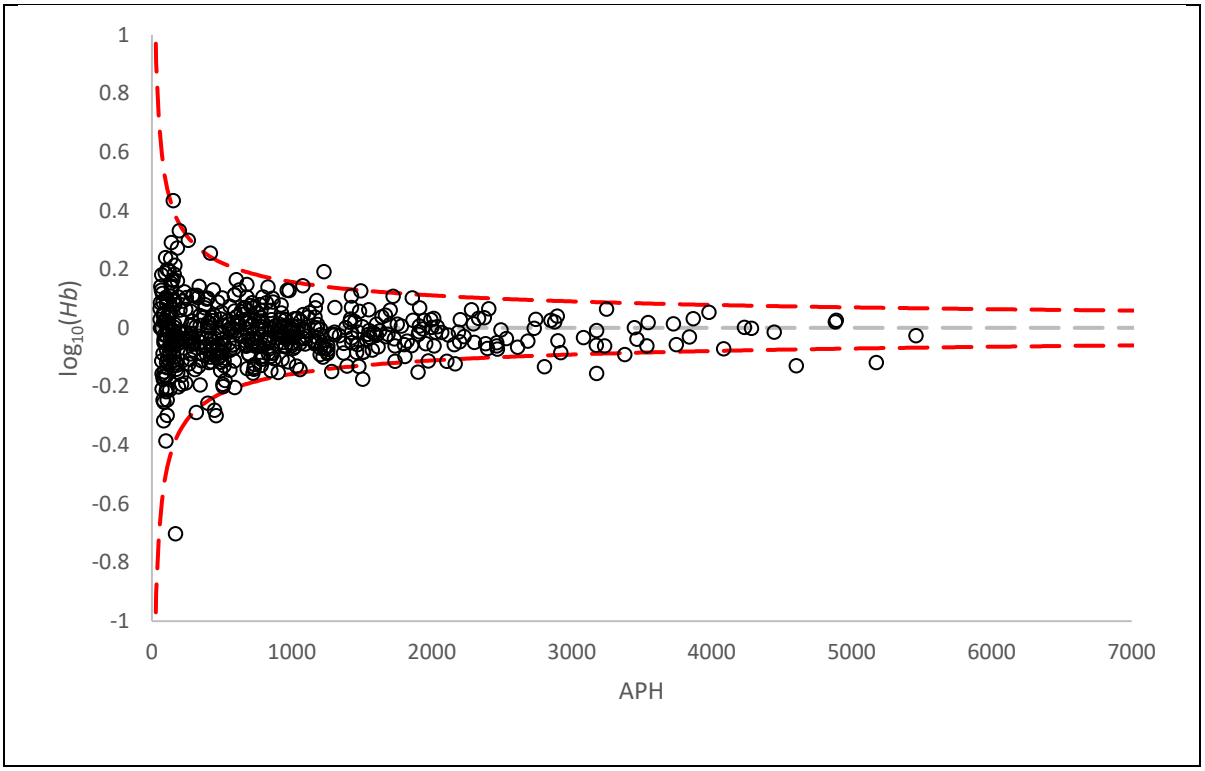


Figure 3: Plot of log(Hb) versus APH for single-source profiles from the combined ISP Laboratory Model Maker dataset. The red dashed lines in each plot indicate the 95% bounds.

Kit settings

The recommended STRmix™ V2.9.1 default parameters for the interpretation of ISP's PP16 profiles run on a 3130 CE instrument and injected using the 3, 5, or 10 second protocols are given in Figure 4 – 6.

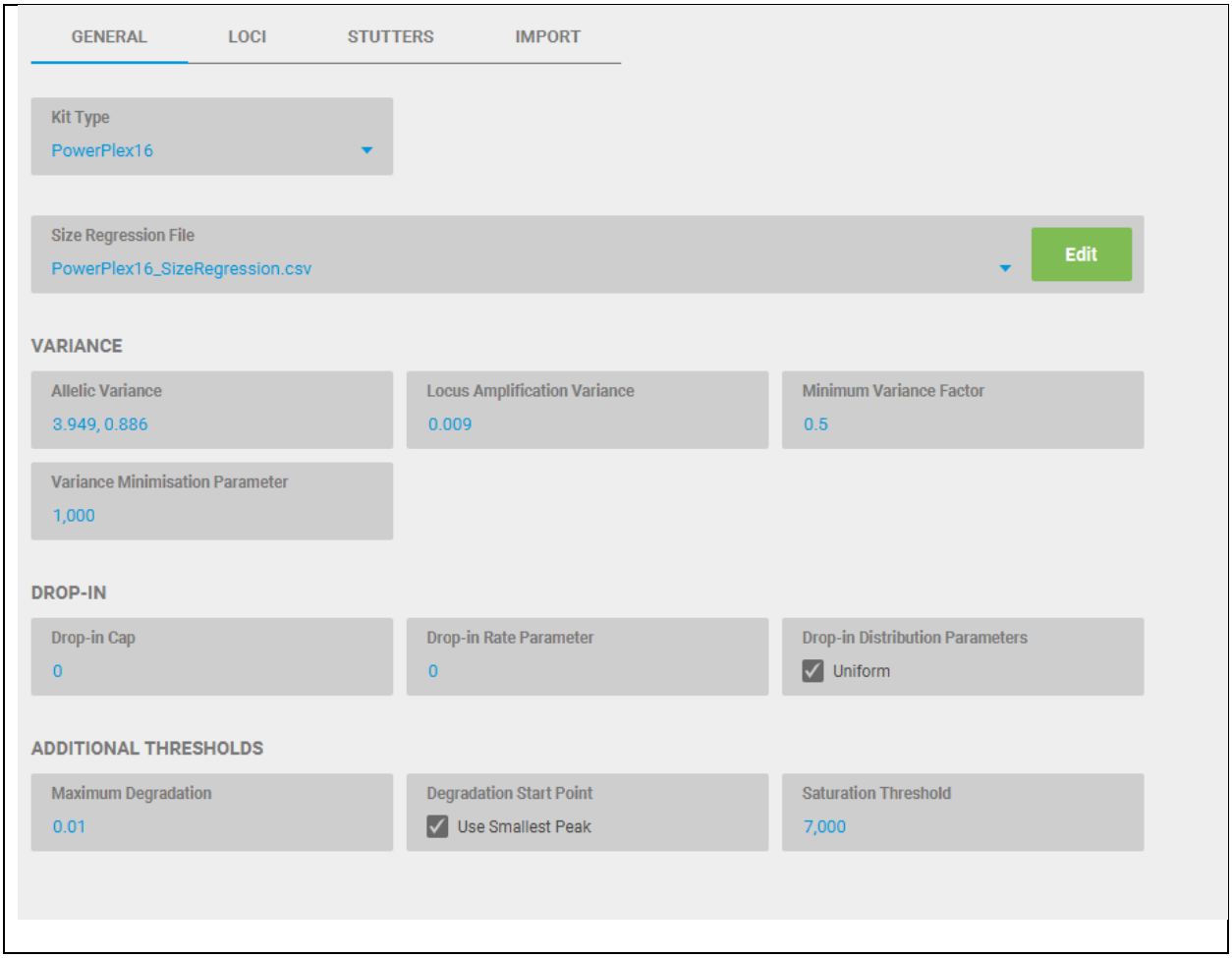


Figure 4: STRmix™ kit settings for PP16 profiles separated on a 3130 CE instrument within the ISP Laboratory. General kit settings shown.

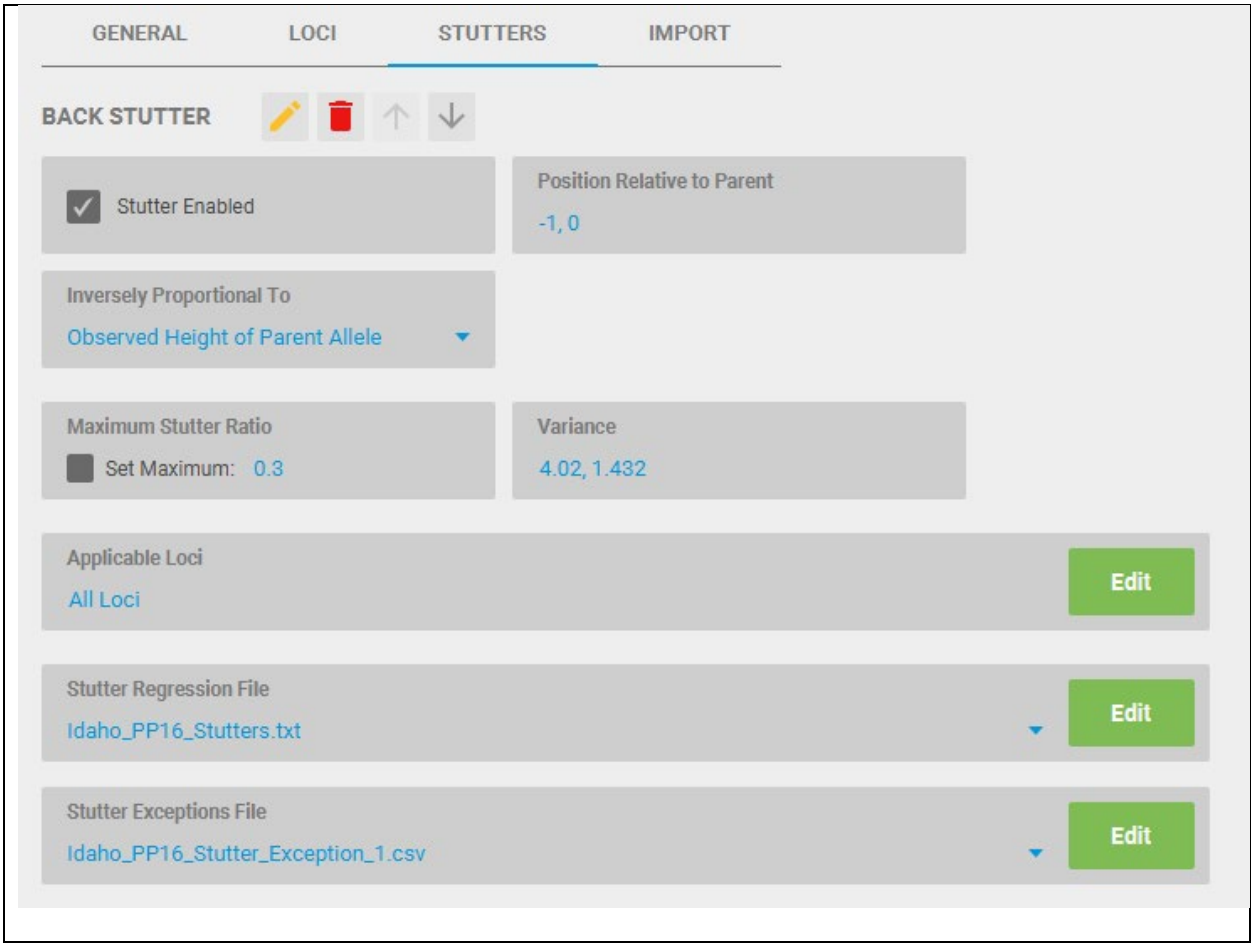


Figure 5: STRmix™ kit settings for PP16 profiles separated on a 3130 CE instrument within the ISP Laboratory. Back stutter settings shown.

LOCUS NAME	GENDER?	REPEAT LENGTH	IGNORE?	DETECTION THRESHOLI
D3S1358	☐	4	☐	50
TH01	☐	4	☐	50
D21S11	☐	4	☐	50
D18S51	☐	4	☐	50
Penta E	☐	5	☐	50
D5S818	☐	4	☐	50
D13S317	☐	4	☐	50
D7S820	☐	4	☐	50
D16S539	☐	4	☐	50
CSF1PO	☐	4	☐	50
Penta D	☐	5	☐	50
AMEL	☑			
vWA	☐	4	☐	50
D8S1179	☐	4	☐	50
TPOX	☐	4	☐	50
FGA	☐	4	☐	50

Figure 6: STRmix™ kit settings for PP16 profiles separated on a 3130 CE instrument within the ISP Laboratory. Locus settings shown

Performance check

To demonstrate the suitability of STRmix™ V2.9.1 and the upgraded kit parameters for the interpretation of PP16 profiles generated within the ISP laboratory, performance checks were undertaken investigating the behavior of the LR using a range of mixed DNA profiles. This includes a review of sensitivity and specificity of the STRmix™ V2.9.1 kit on data generated using the 5 and 10 second injection protocols. The diagnostics from each of the STRmix™ deconvolutions were also reviewed. A comparison of LRs assigned following deconvolution in STRmx™ V2.3.7 and V2.9.1 was also undertaken for the 10 second data set.

Sensitivity and specificity

Two sets of mixtures from the original PP16 STRmix™ validation were used to demonstrate sensitivity and specificity of the STRmix™ V2.9.1 PP16 kit. These mixture sets comprised 56 profiles injected using the 5 second protocol and 58 profiles injected using the 10 second protocol. Both mixture sets were made up of 2, 3, and 4 contributor mixtures that varied in template amount and complexity (relative mixture proportion (M_x) between each contributor). The contributors include homozygote and heterozygote alleles and there is varying amounts of allele sharing across the different loci (SWGDAM standard 4.1.6.5 [3]). Given the template amounts, allele and/or locus drop out was expected to occur within the profiles containing the lower DNA amounts (SWGDAM standard 4.1.7.1 [3]). A list of the mixtures studied and the assigned number of contributors is given in Appendix 1.

The mixtures were re-deconvoluted using the STRmix™ V2.9.1 ISP PP16 kit and LRs were assigned through comparison to a database containing the DNA profiles of the known contributors to the mixture data set and 200+ non-contributors (the same database used within the internal validation of STRmix™ V2.3). An *LR* was assigned considering each of the database individuals in turn as a 'POI' using the Database Search function within STRmix™.

To enable comparison of LRs assigned in each version the sub-sub-source, product rule LRs were assigned using the ISP.PP16.NIST.Cauc.csv allele frequencies. The propositions considered were:

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

H_d: The DNA originated from *N* unknown unrelated individuals

Similar to Taylor [4], sensitivity and specificity data was plotted for each set of data however, average peak height rather than PCR template was used as a variable. These plots are displayed in Figure 7 and Figure 8. Exclusions (*LR*=0) are plotted as log(*LR*)=-40. APH was calculated using unmasked, unshared and non-stutter affected alleles for each contributor in the mixed profile. Where the contributor had completely dropped out of the mixture, an APH of 25 rfu which is half the analytical threshold was applied. The per contributor amount of DNA for H_d true contributors is taken as the lowest APH of the known contributors.

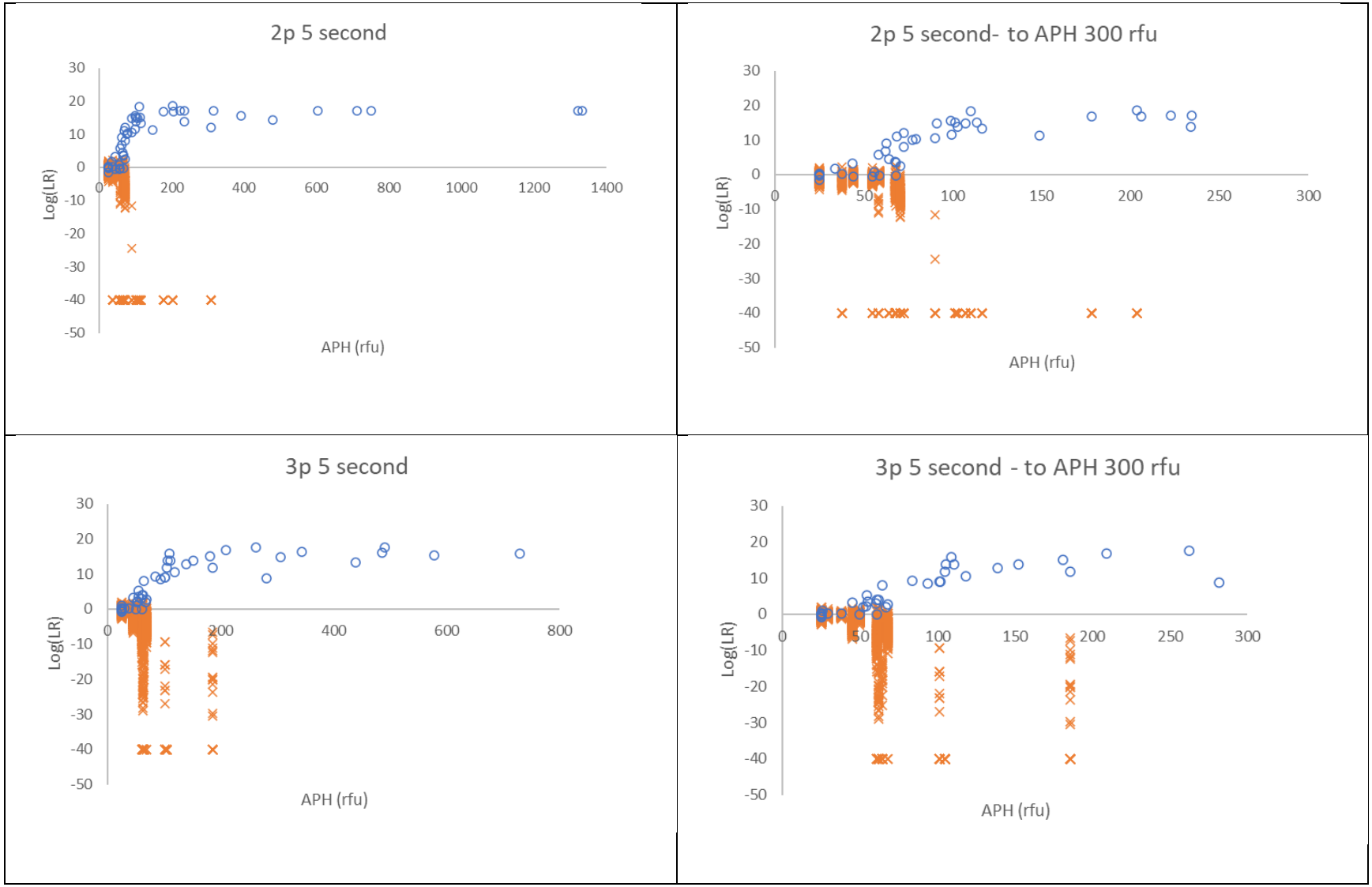


Figure 7: Sensitivity and specificity plots for the 2, 3, and 4 person mixtures (using the 5 second injection protocol). The log(LR)s assigned considering each individual on a database are plotted against APH (rfu). Blue circle datapoints represent the log(LR)s of the known donors and the orange cross datapoints represent the log(LR)s assigned considering non-donors. Two plots are provided for each mixture set, the plot on the left displays the full APH range whilst the plot on the right to 300 rfu only.

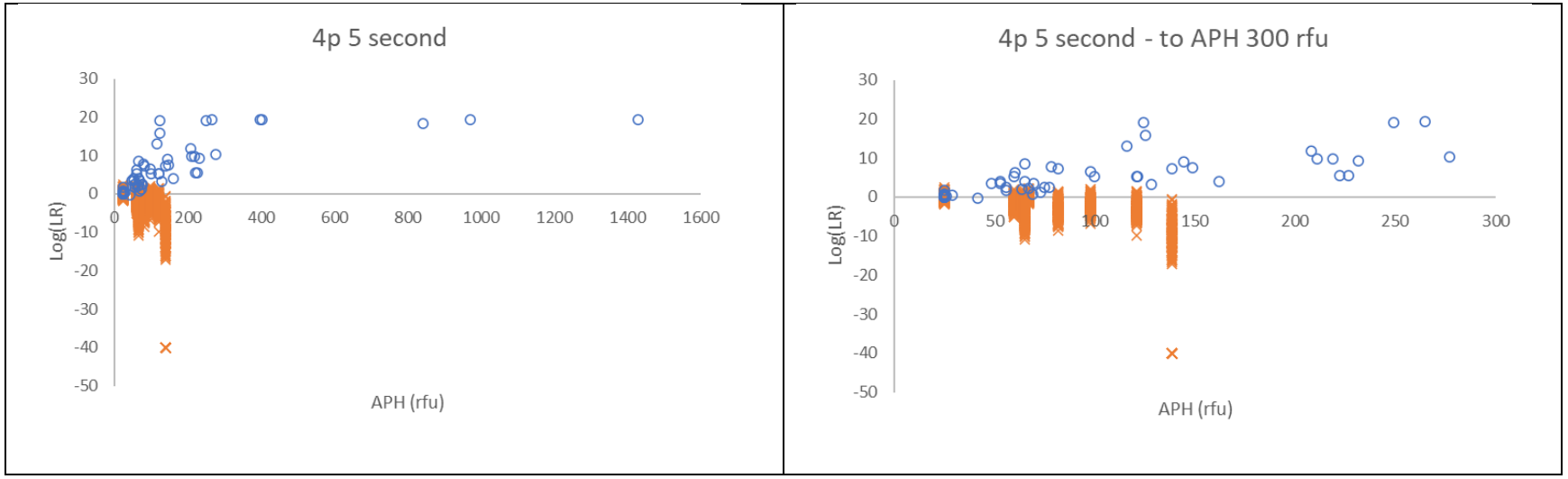


Figure 7 continued

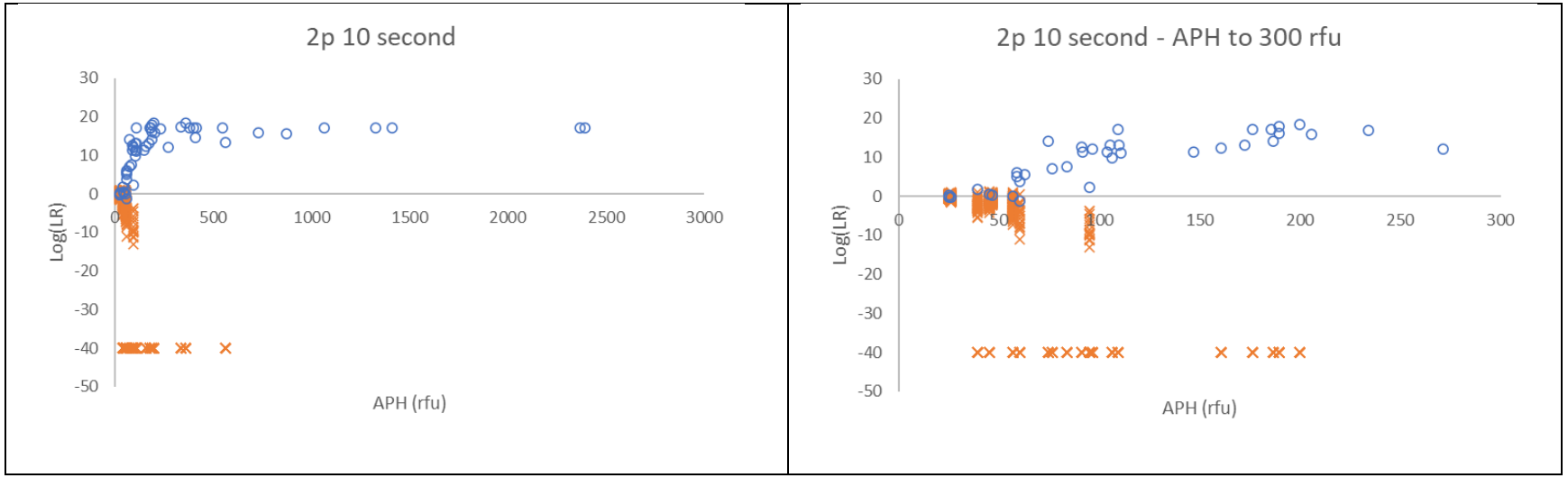


Figure 8: Sensitivity and specificity plots for the 2, 3, and 4 person mixtures (using the 10 second injection protocol). The log(LR)s assigned considering each individual on a database are plotted against APH (rfu). Blue circle datapoints represent the log(LR)s of the known donors and the orange cross datapoints represent the log(LR)s assigned considering non-donors. Two plots are provided for each mixture set, the plot on the left displays the full APH range whilst the plot on the right to 300 rfu only (4p to 500 rfu).

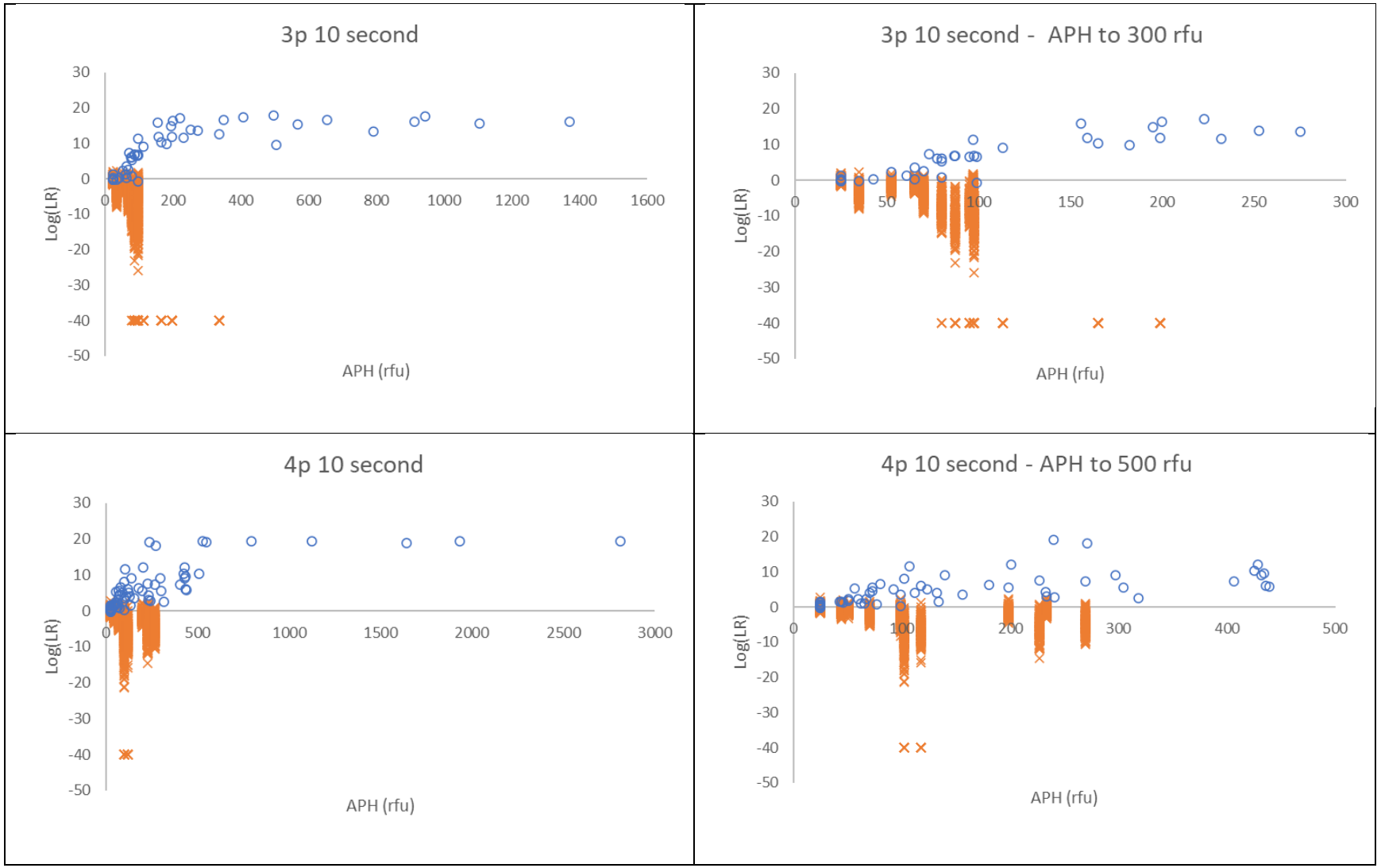


Figure 8 continued

Inspection of the plots in Figure 7 and 8 shows that in general, as average peak height increases the log(*LR*) distributions for H_p true and H_d true are very well separated for two, three and four person mixtures. As the number of contributors increases and the average peak height decreases the two distributions converge on log(*LR*) = 0. At high average peak height, in each mixture class, STRmix™ correctly and reliably gives a high *LR* for true contributors and a low *LR* for non-contributors. At low template and/or high contributor number STRmix™ correctly and reliably reports that the analysis of the sample tends towards uninformative or inconclusive.

In each dataset, a limited number of H_p true data points may be seen below log(LR) = 0 (LR = 1). Each of these were reviewed further and it was found that one or more of the contributors had low peak heights (APH<100rfu) indicating that stochastic sampling effects may be influencing the observed profile. This can lead to less weight being attributed to genotypes of true contributor(s).

Across the datasets, the highest log(LR) assigned to a non-contributor was 2.88 (LR = 772). This was assigned when Random 190 was considered in relation to mixture 1_1_1_1_100pg.2.fsa (10 second injection protocol). Review of this input file revealed that profile is low level (maximum peak height 225 rfu) and partial with only 15 autosomal alleles present meaning that a range of genotypes were accepted at each locus including genotypes with dropout and double dropout. The acceptance of many genotype combinations can lead to adventitious matching of non-contributors. When the Random 190 profile was compared to the mixture it was found that 46% of this non-contributor's alleles were present within the mixed DNA profile with dropout being a viable option at loci where the non-contributor's alleles were not present. Conversely, only 3 of the 15 autosomal peaks in the mixture were not present in the Random 190 profile. The magnitude of the LR is not unexpected given the low-level partial nature of the mixed DNA profile and the assumption of four contributors (the experimentally designed number of contributors). It is recommended that careful consideration as to whether or not to progress an interpretation is given to low level complex profiles where there may be an increased risk of adventitious matching.

Review of Run Diagnostics

STRmix™ includes a number of diagnostics within its reports. These have been deliberately included to assist the user when evaluating the reliability of an interpretation. These may be conveniently categorised into 'primary' and 'secondary' diagnostics. Primary diagnostics include the mixture proportions, genotype weights, and locus LRs. Secondary diagnostics include the average log(likelihood), the Gelman-Rubin convergence diagnostic, and the posterior mean variance parameters. In instances where non-intuitive primary diagnostics are observed, the STRmix™ results should be closely scrutinised however elevated secondary diagnostics do not necessarily invalidate an interpretation. Provided that the primary diagnostics are intuitive, the results are likely still reliable. The secondary diagnostics reported by STRmix™ following interpretation of the mixtures described in Appendix 1 were examined and are discussed further in Appendix 2. For further detail regarding the STRmix™ diagnostics, please refer to [5].

STRmix™ V2.3.7 to V2.9.1 LR comparison

In addition to the sensitivity and specificity plots, a comparison of the log(LR)s for the 2 and 3 contributor mixtures (10 second injection protocol) assigned following deconvolution in STRmix™ V2.3.7, and V2.9.1 was undertaken. This data is displayed below in Figure 9.

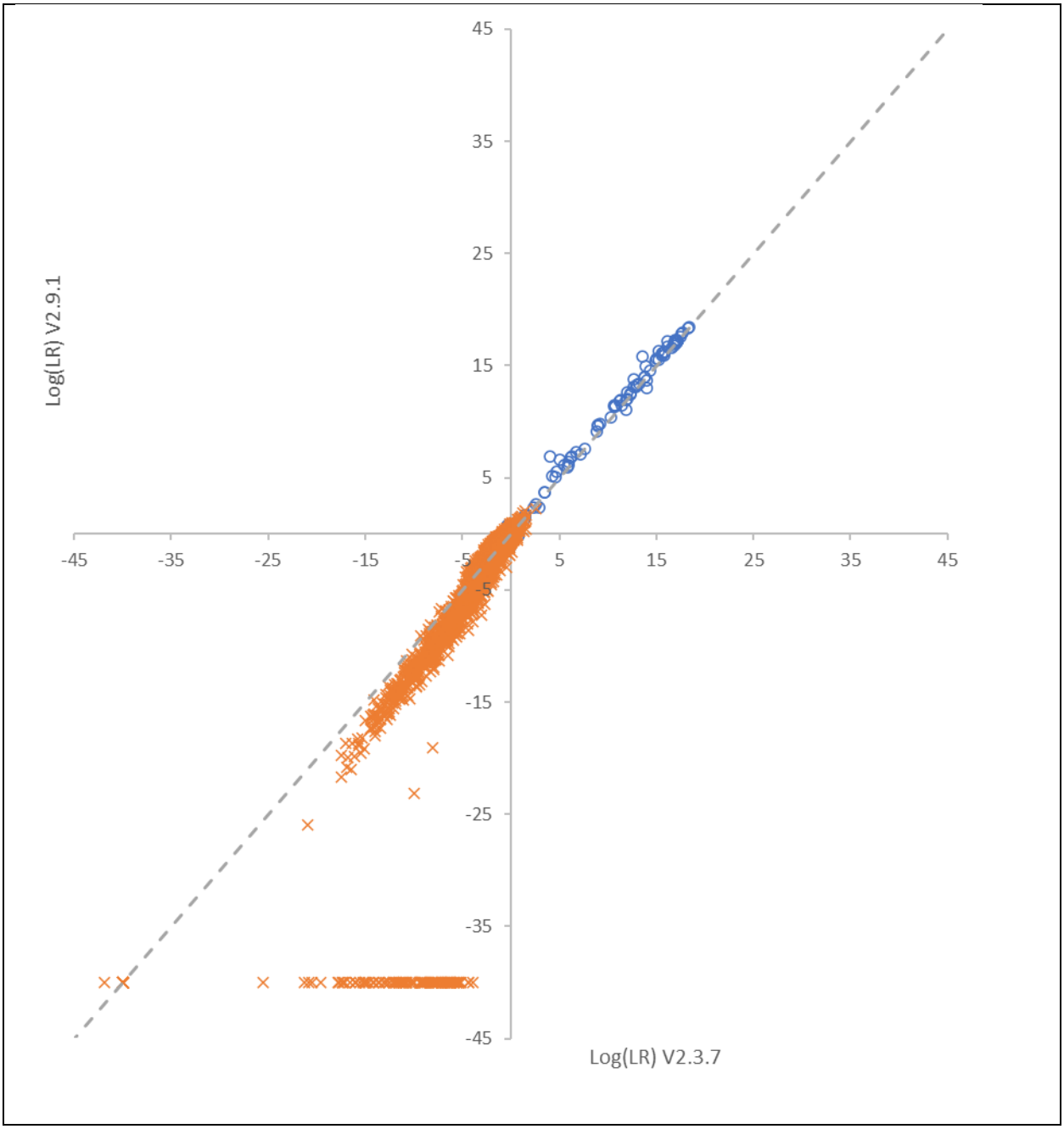


Figure 9: Comparison of log(LR)s assigned following deconvolution in STRmix™ V2.3.7 and V2.9.1. Blue circle datapoints represent known donors whilst orange crosses represent non-contributors. An x = y line of equality has been added to aid visualization.

Due to differences in modelling between STRmix™ V2.3.7 and V2.9.1, and due to small differences between the peak variance parameters used in each ISP PP16 kit, differences in the log(LR)s assigned are expected. Review of this data shows that over 89% of the log(LR)s assigned considering known donors and using STRmix™ V2.9.1 are within one order of magnitude of the log(LR) assigned using STRmix™ V2.3.7. The largest difference was 2.82 orders of magnitude (log(LR) assigned in V2.3.7 = 4.04, log(LR) assigned in V2.9.1 = 6.86) when known donor 'Vin 201' was considered in relation to mixture '1_1_1_100pg.2.fsa'. This profile was low level and partial with evidence of stochastic sampling effects. For example, information at the lower molecular weight D3S1358 locus is not present whilst information is present at higher molecular weight loci. The low-level nature of this profile naturally leads to some variability in the weights between runs and this is likely compounded by the differences in modelling between the versions of STRmix™ in particular the introduction of the LSAE variance parameter. This may have led to some less likely genotype sets not being accepted in the STRmix™ V2.9.1 deconvolution meaning the weights are spread over fewer genotype sets and the LR assigned considering the known donor is higher relative to the LR assigned following deconvolution in STRmix™ V2.3.7. The second largest difference, 2.26 (log(LR) assigned in V2.3.7 = 13.52, log(LR) assigned in V2.9.1 = 15.73) was observed when 'Vin 201' was considered in relation to mixture 3_2_1_100pg.2.fsa. LR from previous runs were undertaken with both the original deconvolution and the STRmix™ V2.9.1 deconvolution and the per locus LRs were compared. The largest differences in the LRs is observed at the D18S51 and Penta E loci. In the original deconvolution, STRmix™ accepted a wide range of genotype combinations in the contributor one position (the position where the highest LR was assigned when 'Vin 201' was considered) at each of these loci diffusing the weights. Whilst in the STRmix™ V2.9.1 deconvolution, fewer genotype combinations were accepted and more weight was focussed on the genotype of the known donor in the contributor 1 position. This is again likely due to the changes in modelling between the different versions of the STRmix™ software, particularly the LSAE variance parameter which may have led to some less likely genotype sets not being accepted.

Whilst no large changes in support for a proposition were noted in the non-donor data points, larger differences between the log(LR)s assigned using each version of STRmix™ may be observed. This is due to the increased variability associated with lower weighted genotype combinations that non-donors tend to align with. Due to general MCMC variability, this degree of difference in log(LR) for non-donors could be expected in two deconvolutions of the mixture in the same version of the software and may not be directly related to STRmix™ version or kit changes. To investigate this, the datapoints that sit at the bottom of the bottom left-hand quadrant of Figure 10 were reviewed further. These datapoints represent a change from exclusionary (V2.3.7) to outright exclusion (V2.9.1). A number of the datapoints could be due to MCMC variation between runs possibly due to non-convergence of the MC chains. Similar to the known donors discussed above, the change in LR value for other datapoints is possibly due to changes in STRmix™ modelling with some unlikely genotype combinations being even less favourable (not accepted) under the constraints of the introduced LSAE variance modelling. A further minor example of a modelling change that impacted a set of non-contributors that had previously (V2.3.7) been assigned a non-zero LR when considered in relation to one of the 2p mixtures (5_1_100pg.2.fsa) was observed. Again, through the use of LR from Previous functionality the locus causing the issue was identified as Penta D and all non-contributors where the exclusionary to exclusion change was observed had the genotype 14,15. When this was investigated further it was found that there is a 12 peak in a stutter position that is exactly 0.3 of

the height of the 13 allele. Despite being relatively tall compared to the 13 peak, STRmix™ V2.3.7 could sometimes account for this peak and any other peaks less than or equal to the back stutter max (0.3) as stutter allowing other combinations such as -1,-1 (which allows for the association of the non-contributors with the genotype 14,15) to be accepted. However, in later versions of the software any peak that is greater than or equal to the back stutter max must be considered to have some allelic component meaning that a double dropout option could not be considered and any contributor that does not have either an 11, 12, or 13 allele (the alleles present within the input file at this locus) is excluded.

In summary, the plots in Figure 7 and 8 demonstrate that at high average peak height STRmix™ correctly and reliably gave high *LR*s for known contributors and a low or exclusionary LR for non-contributors. At low average peak height and higher contributor number profiles STRmix™ correctly and reliably reported that the analysis of the sample tends towards uninformative or inconclusive. The plots also help to inform the limits of STRmix™, particularly the lower limit of DNA where an H_p true hypothesis still results in an *LR* greater than 1 and the limit where false positives may arise (an *LR* greater than 1 where H_d is true). The diagnostic values discussed in Appendix 2 are appropriate given profile features and the comparison of log(LR)s assigned demonstrates little difference in the LRs assigned following deconvolution in each version of the software. These performance check results support the use of the updated STRmix™ V2.9.1 PP16 kit for the interpretation of data generated at the ISP laboratory.

Signatures

Taylor Maichak

ISP Laboratory STRmix™ project lead and Technical Leader

This work has been reviewed and it has been determined that STRmix™ V2.9.1 is suitable for its intended use for interpretation of PowerPlex® 16 crime profiles at Idaho State Police laboratory. The project work has met the validation requirements as required by A2LA and QAS.

ISP Quality Manager

References

1. Bright, J.-A., et al., *Determination of the variables affecting mixed MiniFiler™ DNA profiles.* Forensic Science International: Genetics, 2011. **5**(5): p. 381-385.
2. Bright, J.-A., J. Turkington, and J. Buckleton, *Examination of the variability in mixed DNA profile parameters for the Identifiler(TM) multiplex.* Forensic Sci Int Genet, 2009. **4**.
3. Scientific Working Group on DNA Analysis Methods. *SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems*. 2015; Available from: https://1ecb9588-ea6f-4feb-971a-73265dbf079c.filesusr.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf.
4. Taylor, D., *Using continuous DNA interpretation methods to revisit likelihood ratio behaviour.* Forensic Science International: Genetics, 2014. **11**: p. 144-153.
5. Russell, L., et al., *A guide to results and diagnostics within a STRmix™ report.* WIREs Forensic Science, 2019. **1**(6): p. e1354.

Appendix 1: Summary of profiles evaluated during performance check

Profiles injected using 5 second protocol

Sample ID	Number of contributors used in STRmix™ deconvolution
20_1_50pg.fsa	2
20_1_100pg.fsa	2
20_1_150pg.fsa	2
20_1_200pg.fsa	2
20_1_250pg.fsa	2
1_1_50pg.fsa	2
1_1_100pg.fsa	2
1_1_200pg.fsa	2
1_1_300pg.fsa	2
1_1_400pg.fsa	2
1_1_500pg.fsa	2
3_1_50pg.fsa	2
3_1_100pg.fsa	2
3_1_150pg.fsa	2
3_1_200pg.fsa	2
3_1_250pg.fsa	2
5_1_50pg.fsa	2
5_1_100pg.fsa	2
5_1_150pg.fsa	2
5_1_200pg.fsa	2
5_1_250pg.fsa	2
10_1_50pg.fsa	2
10_1_100pg.fsa	2
10_1_150pg.fsa	2
10_1_200pg.fsa	2
10_1_250pg.fsa	2
1_1_1_100pg.fsa	3
1_1_1_200pg.fsa	3
1_1_1_300pg.fsa	3
1_1_1_400pg.fsa	3
1_1_1_500pg.fsa	3
3_2_1_50pg.fsa	3
3_2_1_100pg.fsa	3
3_2_1_150pg.fsa	3
3_2_1_200pg.fsa	3

3_2_1_250pg.fsa	3
10_5_1_50pg.fsa	3
10_5_1_100pg.fsa	3
10_5_1_150pg.fsa	3
10_5_1_200pg.fsa	3
10_5_1_250pg.fsa	3
10_5_2_1_100pg.fsa	4
10_5_2_1_150pg.fsa	4
10_5_2_1_200pg.fsa	4
10_5_2_1_250pg.fsa	4
1_1_1_1_100pg.fsa	4
1_1_1_1_200pg.fsa	4
1_1_1_1_300pg.fsa	4
1_1_1_1_400pg.fsa	4
1_1_1_1_500pg.fsa	4
4_3_2_1_50pg.fsa	4
4_3_2_1_100pg.fsa	4
4_3_2_1_150pg.fsa	4
4_3_2_1_200pg.fsa	4
4_3_2_1_250pg.fsa	4
10_5_2_1_50pg.fsa	4

Profiles injected using 10 second protocol

Sample ID	Number of contributors used in STRmix™ deconvolution
1_1_50pg.2.fsa	2
1_1_100pg.2.fsa	2
1_1_200pg.2.fsa	2
1_1_300pg.2.fsa	2
1_1_400pg.2.fsa	2
1_1_500pg.2.fsa	2
3_1_50pg.2.fsa	2
3_1_100pg.2.fsa	2
3_1_150pg.2.fsa	2
3_1_200pg.2.fsa	2
3_1_250pg.2.fsa	2
5_1_50pg.2.fsa	2
5_1_100pg.2.fsa	2
5_1_150pg.2.fsa	2
5_1_200pg.2.fsa	2
5_1_250pg.2.fsa	2
10_1_50pg.2.fsa	2
10_1_100pg.2.fsa	2
10_1_150pg.2.fsa	2
10_1_200pg.2.fsa	2
10_1_250pg.2.fsa	2
20_1_50pg.2.fsa	2
20_1_100pg.2.fsa	2
20_1_150pg.2.fsa	2
20_1_200pg.2.fsa	2
20_1_250pg.2.fsa	2
1_1_1_100pg.2.fsa	3
1_1_1_200pg.2.fsa	3
1_1_1_300pg.2.fsa	3
1_1_1_400pg.2.fsa	3
1_1_1_500pg.2.fsa	3
3_2_1_50pg.2.fsa	3
3_2_1_100pg.2.fsa	3
3_2_1_150pg.2.fsa	3
3_2_1_200pg.2.fsa	3
3_2_1_250pg.2.fsa	3
10_5_1_50pg.2.fsa	3

10_5_1_100pg.2.fsa	3
10_5_1_150pg.2.fsa	3
10_5_1_200pg.2.fsa	3
10_5_1_250pg.2.fsa	3
1_1_1_50pg.2.fsa	3
4_3_2_1_50pg.2.fsa	4
4_3_2_1_100pg.2.fsa	4
4_3_2_1_150pg.2.fsa	4
4_3_2_1_200pg.2.fsa	4
4_3_2_1_250pg.2.fsa	4
10_5_2_1_50pg.2.fsa	4
10_5_2_1_100pg.2.fsa	4
10_5_2_1_150pg.2.fsa	4
10_5_2_1_200pg.2.fsa	4
10_5_2_1_250pg.2.fsa	4
1_1_1_1_50pg.2.fsa	4
1_1_1_1_100pg.2.fsa	4
1_1_1_1_200pg.2.fsa	4
1_1_1_1_300pg.2.fsa	4
1_1_1_1_400pg.2.fsa	4
1_1_1_1_500pg.2.fsa	4

Appendix 2: Review of Secondary Run Diagnostics

Secondary diagnostics are a useful guide to provide confidence the STRmix™ interpretation has progressed as expected. Individual secondary diagnostics may indicate whether a more comprehensive review is warranted, however analysts should not rely on these diagnostics alone. Elevated values for one of these diagnostics may not necessarily mean the results are unfit for purpose. To put in context the range of diagnostic values that can be expected from ISP PP16 data, a discussion of the secondary run diagnostics obtained from STRmix™ V2.9.1 deconvolutions of the mixtures studied as part of this performance check is provided below.

Effective sample size (ESS)

This is a measure of the degree of correlation within the accepts of a STRmix™ deconvolution. It is used within the Highest Posterior Density (HPD) method to help take into account uncertainty in the weights. ESS is used to convert the full sample set of iterations that includes many with correlation into a set of independent samples that may be resampled from during the HPD process. If a sample set for a chain is fully correlated then an ESS value of 1 would be observed and this indicates a problem with the deconvolution. As there are 8 chains, full correlation across each of these chains would display as 8. No ESS values of 8 were observed within this limited data set but analysts are advised to review this as part of their STRmix™ output and interpretation process. It has been demonstrated that correlation can be seed (starting point) related, so if a value of 8 is observed simply rerunning the deconvolution with a different seed is recommended.

Average log(likelihood)

STRmix™ uses a biological model to generate an expected DNA profile which is then compared with the observed profile. When assessing the fit of the expected profile with the observed, STRmix™ calculates a 'grade', referred to as a log(likelihood). The average log(likelihood) diagnostic reported in the STRmix™ output is the average of the log(likelihood) values across all post-burn-in iterations. The larger this value is, the better STRmix™ has been able to describe the observed data. A low or negative value suggests that STRmix™ has not been able to describe the data very well given the information it has been provided with. Reasons why this value may be low or negative include:

1. The profile is simply low level and there is very little data making up the likelihood,

2. There are large stochastic events in the STRmix™ run (e.g., large heterozygote peak imbalances or variation in mixture proportions across the profile). These may be forced by mis-assignment of the number of contributors,

3. Data has been removed that was real, in particular stutter peaks, and must now be described within STRmix™ by dropout, and

4. Artefactual peaks have been left labelled and must now be accounted for within STRmix™ by e.g., drop-in.

As per point 1 above, it is important to note that low or negative average log(likelihood) values may legitimately be produced when interpreting low level DNA profiles. As such, low or negative average log(likelihood) values do not necessarily indicate that the STRmix™ results are unreliable.

The average log(likelihood) diagnostic for each of the interpretations from the 5 and 10 second injection protocol datasets is plotted against the NOC assigned in the original validation in below.

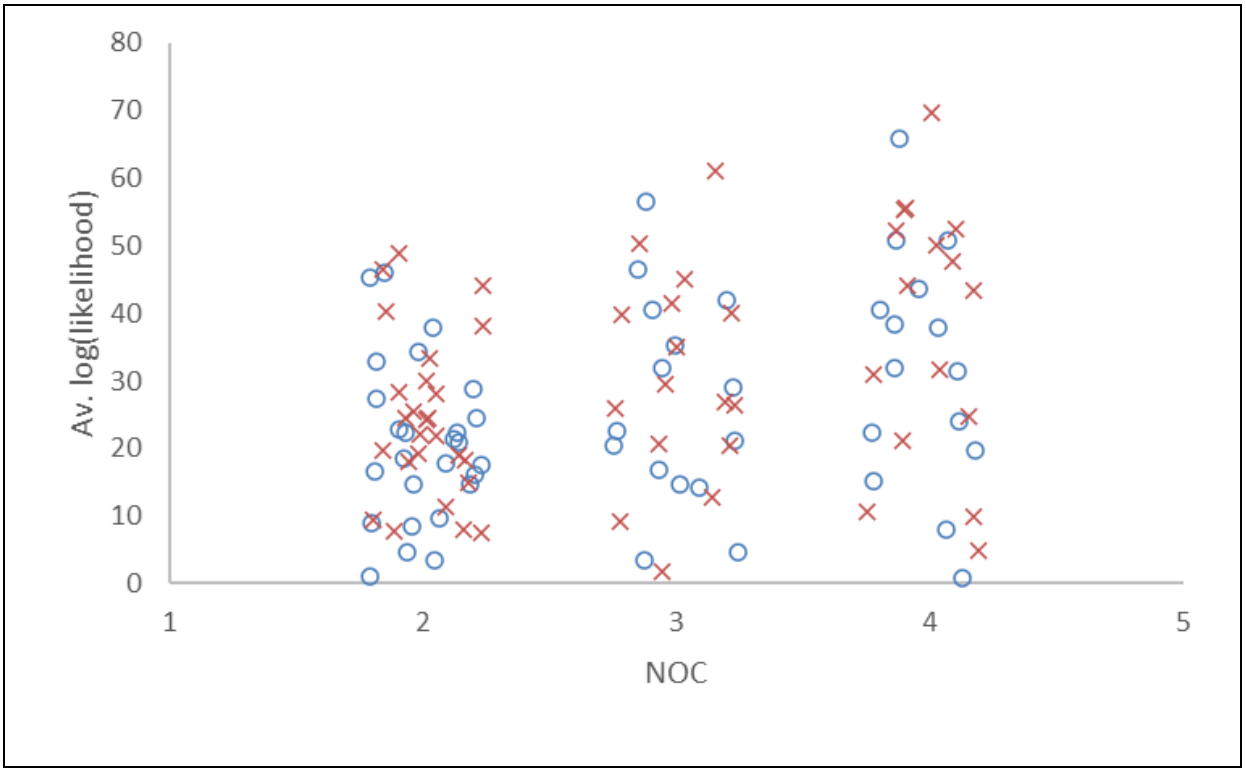


Figure 10: Plot of average log(likelihood) diagnostic versus assigned number of contributors. Blue circle datapoints represent the 5 second injection protocol dataset and the red crosses the 10 second injection protocol dataset

The smallest value observed was approximately 0.879 and was recorded for a low-level mixture (1_1_1_1_100pg.fsa, 5 second injection) deconvoluted under the assumption it originates from four contributors. Review of the input file revealed that this profile had only two autosomal peaks present. All other profiles that resulted in an average log(likelihood) less than 5 were reviewed and found to have 8 or fewer peaks present with heights less than 350 rfu.

The largest value observed was approximately 69.60 and was recorded for a four-person mixture with strong peak heights (10_5_2_1_200pg.2.fsa), 10 second injection)

Gelman-Rubin convergence diagnostic

Ideally, each MCMC chain will reach the area of high probability within the sample space during burn-in and will continue to sample from this space during the post-burn-in MCMC. This is referred to as 'convergence'. If the chains spend their time in different spaces during the post-burn-in MCMC then it is

likely that the analysis has not been run for long enough. The Gelman-Rubin (GR) convergence diagnostic included in the STRmix™ report can indicate to the user if the Markov chains have not sufficiently converged. If the chains have been sampling from the same space, then the GR diagnostic should be close to 1.0. Notionally, values above 1.2 indicate that the chains may not be nearing convergence. It is important to note that the GR diagnostic output by STRmix™ is a summary statistic: values less than 1.2 do not guarantee that all parameters have converged whilst values greater than 1.2 do not necessarily indicate that the results are unreliable.

In rare instances, one (or more) chain(s) may fail to find the area of high probability space altogether. This is referred to as a wandering chain and typically leads to substantially elevated GR diagnostics. Often, the genotypes accepted at one or more loci will not be intuitive in instances where there has been a wandering chain. Simply re-running the interpretation will typically recover the GR and produce sensible results. However, not all causes of an elevated GR can be addressed in this way, therefore as with all run diagnostics it is recommended that both the input file and the primary and secondary outputs of runs with excessive values are closely scrutinized.

The GR convergence diagnostic for each of the deconvolutions is plotted against assigned NOC in Figure 11 below.

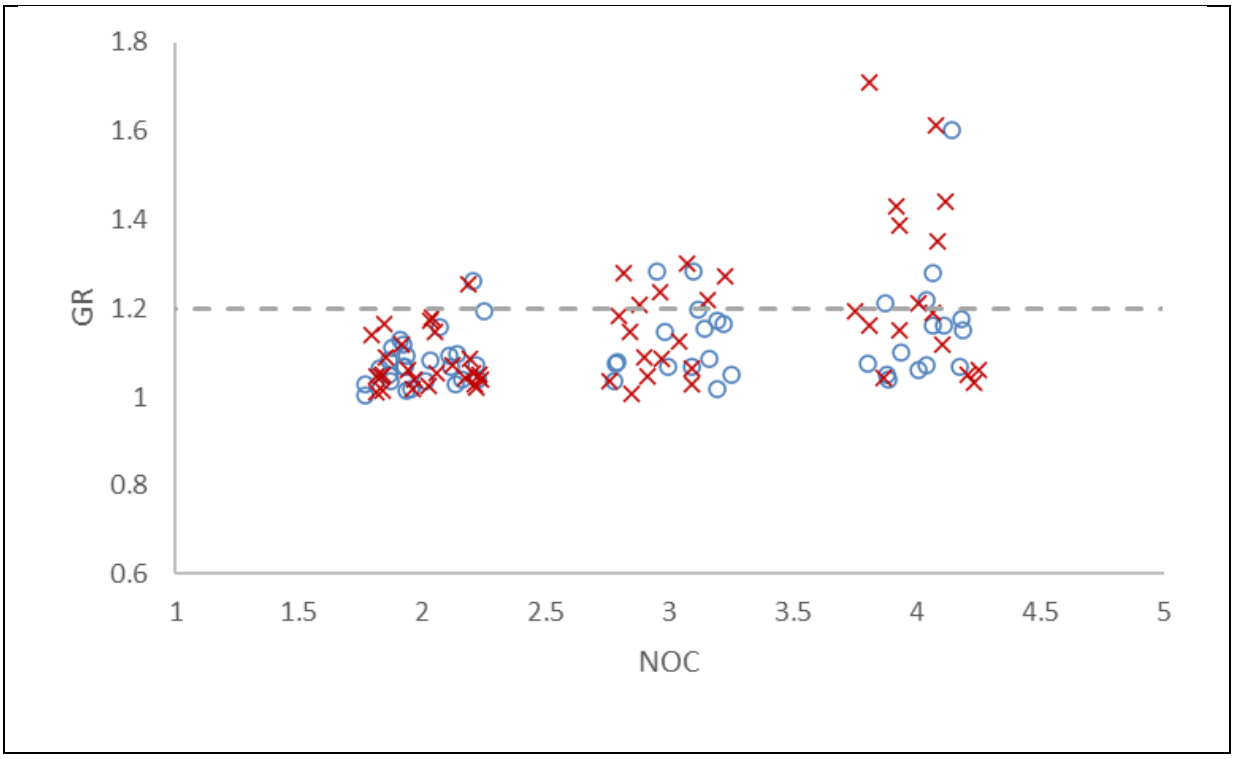


Figure 11: Plot of Gelman-Rubin (GR) convergence diagnostic versus number of contributors. The dashed line indicates a GR value of 1.2. Blue circle datapoints represent the 5 second injection protocol dataset and the red crosses the 10 second injection protocol dataset.

Over 87% of the mixtures of the 5 second dataset and 75% of the 10 second dataset resulted in GRs less than 1.2. The remaining mixtures resulted in mildly elevated GRs greater than 1.2 but not excessive. The highest GR (1.71) was observed following the deconvolution of a four-person mixture

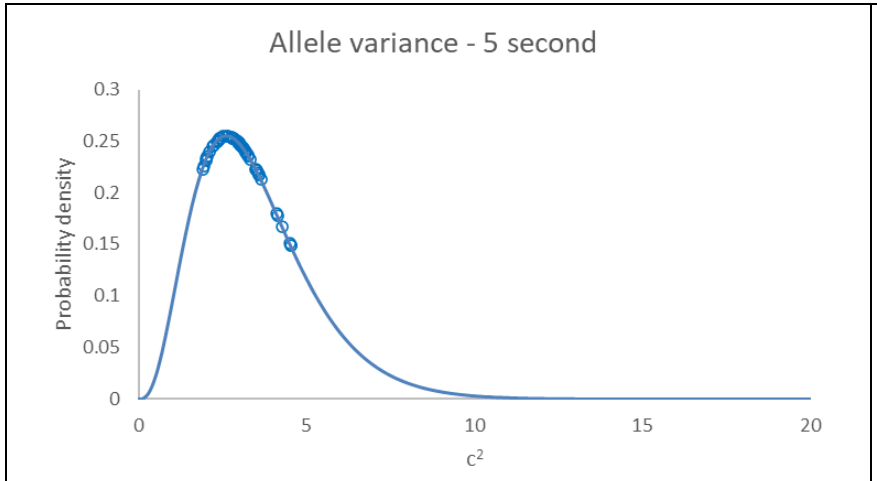
(4_3_2_1_250pg.2.fsa, 10 second dataset). The mixture design makes this a naturally complex profile however, there also appears to be an increase in the blue peak heights at the higher molecular weight end of the profile making this an even more challenging profile for STRmix™ to deconvolute. This increase in peak height across a dye is counter to the STRmix™ model that generally assumes a decrease in peak height from low molecular weight loci to the higher molecular weight loci. The LSAE variance value for this deconvolution is also elevated indicating that STRmix™ likely used higher LSAE values and hence higher LSAE variance values during the post-burnin phase to accommodate these taller peaks. Despite the mildly elevated GR, the LRs assigned considering true donors and non-donors were intuitive for this sample. The other deconvolutions resulting in GRs larger than 1.2 were all relatively complex interpretations; high order mixtures (assigned number of contributors), low template or mixtures where two or more contributors are in close proportions with relatively low peak heights. As described in the original validation document, complex profiles or profiles where a GR in excess of 1.2 has been observed can be deconvoluted with an increased numbers of accepts to allow the chains to converge on the same high probability space.

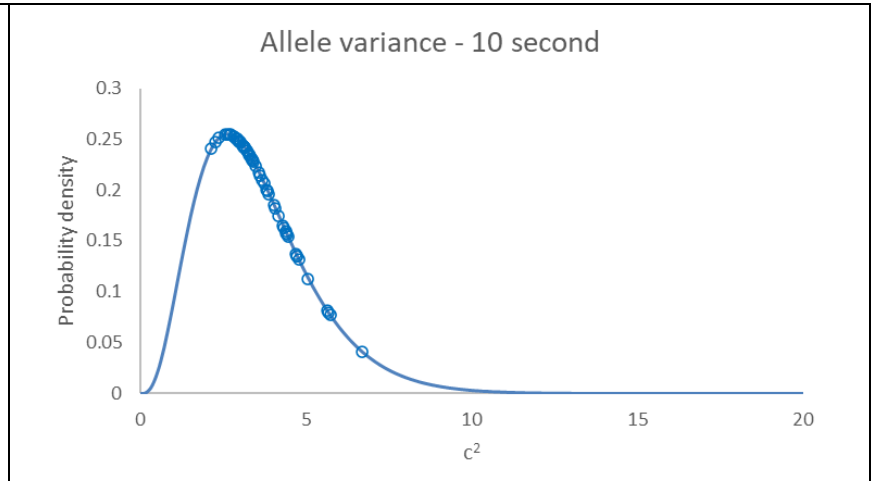
Posterior variance parameters

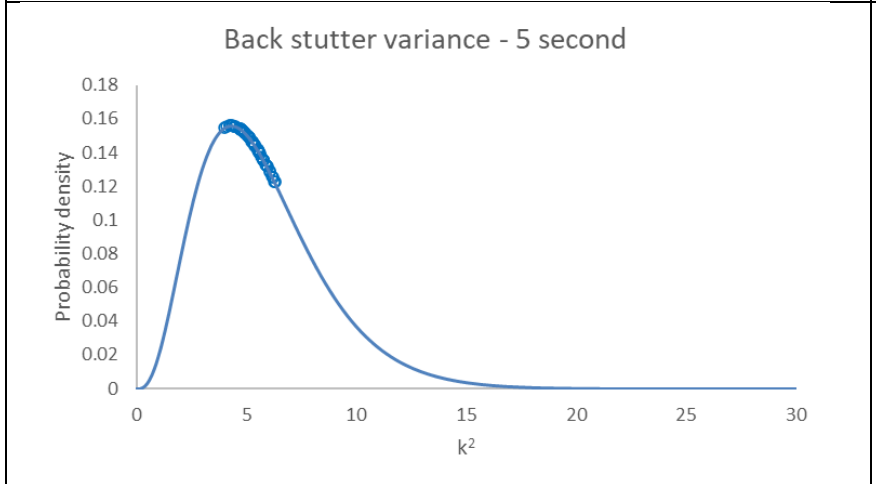
Within the STRmix™ report, the posterior mean variance parameters are overlaid on the relevant prior distributions. Ideally, each of the posterior variance parameters should sit within the body of the relevant prior distribution. Values that fall in the right-hand tail of the prior distribution may warrant further investigation. A large allele variance parameter in conjunction with a low or negative average log(likelihood) diagnostic may indicate that the number of contributors to the profile has been mis-assigned. Excessive stutter variance parameters may be due to the inadvertent application of a stutter filter during CE profile analysis. As with the other secondary diagnostics described above, elevated variance parameters do not necessarily invalidate the results. Provided that the primary diagnostics are intuitive, the STRmix™ results are likely reliable.

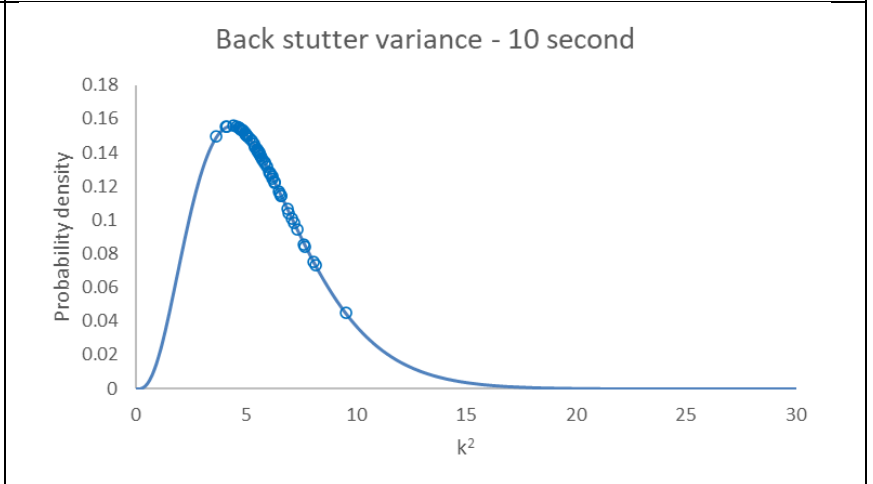
The posterior variance parameters for each of the mixed DNA profiles interpreted in STRmix V2.9.1 along with their prior distributions are provided in Figure 12 below. Most of the values reported by STRmix™ were acceptable and did not warrant further investigation. The limited number of outliers observed are discussed below.

Three of the 5 second deconvolutions and seven of the 10 second led to LSAE values greater than 0.02. These were each reviewed further and found to be due either to low peak heights and stochastic sampling effects or due to higher peak heights at the higher molecular weight loci of the blue dye (as described in the Gelman-Rubin section above). The weights output for profiles such as these should be considered carefully and where possible rework such as reamplification should be attempted to address any issues.


Allele variance - 5 second
Probability density
0.3
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
20
c²


Allele variance - 10 second
Probability density
0.3
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
20
c²


Back stutter variance - 5 second
Probability density
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
5
10
15
20
25
30
k²


Back stutter variance - 10 second
Probability density
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
5
10
15
20
25
30
k²

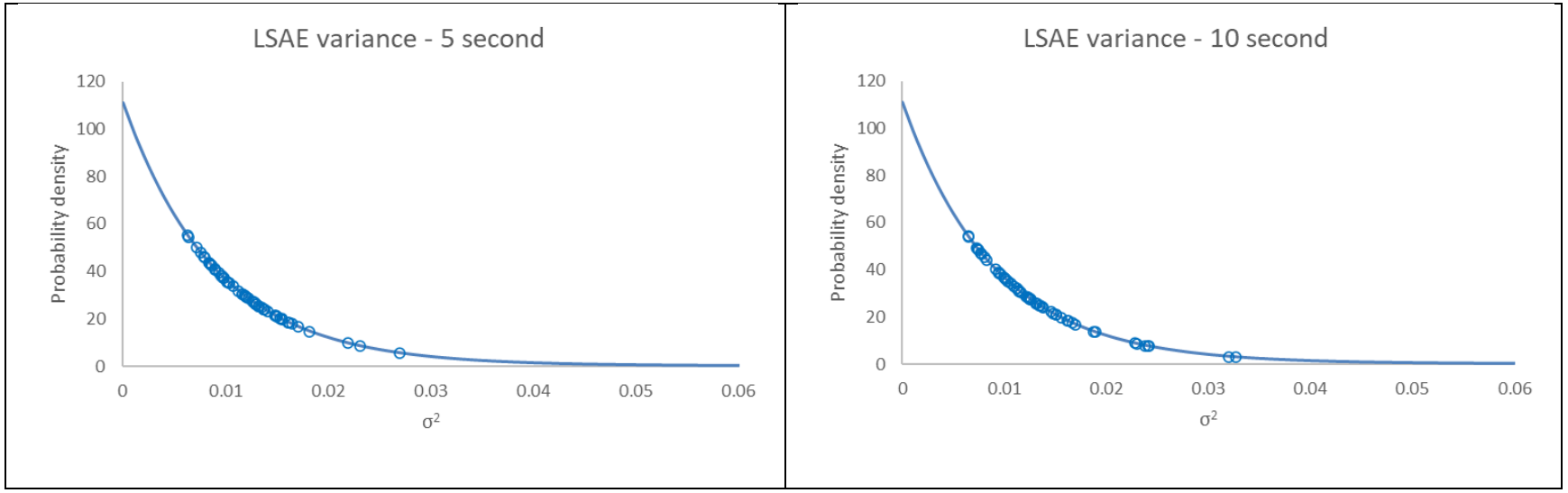


Figure 12: Plots of peak height and LSAE variance. The prior distributions for each parameter are plotted in blue. The posterior mean variance parameters are plotted as blue circle data points. These are the optimised values reported by STRmix™ following profile interpretation.