2.8 Implementation and Validation Guide

As employees of ESR, STRmix™ staff operate within the relevant contractual, legislative and ethical responsibilities. As forensic scientists, we utilise knowledge and experience to enhance the welfare and safety of all members of the community without discrimination or prejudice, treating all with respect, honesty, equality, and integrity. We follow the Code of Professional Practice for members of the Australian and New Zealand Forensic Science Society (ANZFSS). It is understood that by submitting data to ESR for analysis and/or interpretation, customers confirm that these samples were collected with consent, and ESR takes no responsibility for verifying that consent has been obtained.

This document should be read in conjunction with the *STRmix™ V2.8 User's Manual* and the *STRmix™ V2.8 Operation Manual*. The STRmix™ team offers services to assist with implementation and validation activities. Contact support@strmix.com for more information. The Institute of Environmental Science & Research Ltd. (ESR) has taken all reasonable measures to ensure that the information and data presented in this document is accurate and current. However, ESR makes no express or implied warranty regarding such information or data, and hereby expressly disclaims all legal liability and responsibility to persons or entities that use or access this document and its content.

CONTENTS

1 INTRODUCTION

Prior to the internal validation of STRmix™ within a laboratory a number of laboratory-specific parameters are required to be determined. These include the analytical threshold(s) used when analysing raw DNA data following capillary electrophoresis (CE), expected stutter ratios pertaining to the multiplex kit(s) and instrumentation used within the laboratory, the saturation threshold of the CE instrumentation, drop-in parameters, and the variance parameters used within STRmix™ (allele and stutter peak height variance parameters and locus specific amplification efficiency (LSAE) variance). The estimation of each of these parameters is discussed within Section 2 of this manual.

Within Section 3, guidelines for the internal validation of STRmix™ following the recommendations of the Scientific Working Group on DNA Analysis Methods (SWGDAM) are described. These include experiments to investigate the precision, reproducibility, sensitivity, and specificity of the software.

In Section 4, we describe experiments designed to assist with the internal validation of interpretations using PCR replicates, multi-kit interpretations, interpretations using the variable number of contributors function, use of the M_x Priors feature, the Mix to Mix function and the Top Down Approach. The use of these methods is optional within a laboratory and their internal validation is not prescribed within any published guidelines. We have described several recommended experiments based on good practise and on our understanding of the techniques.

2 STRMIX™ IMPLEMENTATION

2.1 Implementation introduction

Laboratory-specific parameters that must be optimised prior to validation include:

1. Analytical/detection threshold(s),
2. Expected stutter ratios,
3. The CE instrument's saturation limit,
4. Drop-in parameters,
5. Allele and stutter peak height variance prior distributions,
6. LSAE variance prior distribution, and
7. Parameters for all relevant populations including allele frequencies and theta (F_{ST}) value.

It is probable that different amplification protocols, CE platforms, and CE injection protocols will have different parameters for some if not all of these. Some of these variables, for example stutter ratios [1] have been demonstrated to not differ significantly between laboratories using the same multiplex kit and PCR protocol. Analytical thresholds, variance prior distributions, and saturation thresholds have also been shown to be comparable between different laboratories under certain conditions, for example, between different laboratories within a wider laboratory system or laboratories using the same kit, PCR protocol, and instrumentation [2].

The determination of each of these parameters is discussed in turn in the following sections within this manual. It is recommended that you review this manual with a member of the STRmix™ team prior to the commencement of implementation and validation of STRmix™ within your laboratory, to help inform the scope of the activities required.

2.1 Analytical threshold (AT)

The AT or limit of detection is defined as the rfu value where a true DNA signal can reliably be distinguished from instrument noise. There are several published papers describing how the analytical threshold may be determined [3-5]. Typically, a value is set some number of standard deviations above the average observed peak height of baseline instrument noise. This parameter is likely to have already been determined by a laboratory and therefore detailed instructions on how to derive it are not repeated here. Further advice can be found in Taylor et al. [3]. In this study, regression analysis showed small but significant effects of dye colour and CE instrument model on both the mean and standard deviation (sd) of baseline noise. Taylor et al. also observed template to have an effect on baseline noise, highlighting the importance of using data collected from DNA-positive samples in addition to negative controls when estimating AT. The importance of using DNA-positive samples in addition to negative controls was reinforced by Monich et al. [6].

Within STRmix™, the detection threshold parameter (in rfu) is used within the modelling of dropout and drop-in during profile interpretation. When an interpretation is commenced, STRmix™ V2.8 will alert the user to peaks within the input file with height below the detection threshold(s) input within the STRmix™ kit settings. It is

recommended that such peaks are ignored during interpretation. Per dye or per locus thresholds can be set within STRmix™, if required. For casework application, the detection thresholds set within STRmix™ must align with the AT values used to analyse the casework profiles.

2.2 Stutter

In general, stutter ratios are calculated by:

$$SR = \frac{O_{a\pm(i,j)}}{O_a}$$

Where $O_{a\pm(i,j)}$ is the observed height of the stutter peak, O_a is the observed height of the parent allele, a, and (i, j) is the stutter's position relative to its parent peak. i represents the number of whole STR units the stutter is located from the parent peak while j represents any additional base pairs required to further define the location of the stutter. For example, (-1,0) describes a stutter product that is one whole repeat unit shorter than the parent allele (i.e. back stutter) whereas (0,-2) or (-1,2) describe a stutter product that is two base pairs shorter than the parent allele (i.e. half back stutter).

In STRmix™ V2.6 and later, any type of stutter may be modelled. This is referred to as *generalised stutter modelling*. The following stutter variants are commonly observed in modern forensic multiplex kits:

- Back stutter (one repeat unit shorter than the parent allele),
- Forward stutter (one repeat unit larger than the parent allele),
- Double back stutter (two repeat units shorter than the parent allele), and
- Minus two base pair stutter/half back stutter (2 base pairs shorter than the parent allele and commonly observed at the SE33 and D1S1656 loci).

Expected stutter ratio values are determined by analysing single-source samples amplified using your laboratory's optimum template amount. A minimum of approximately 100 profiles is recommended. These samples should originate from different donors selected to maximise the range of alleles covered at each locus. Stutter ratios for some variants (e.g. double back stutter) have been observed to be relatively low; for such variants, obtaining sufficient empirical data to inform and develop suitable models is not straightforward. We recommend that a reduced AT is used during CE data analysis of these profiles in order to detect more stutter information.

Within STRmix™, expected stutter ratio values are calculated using two types of file: text files (.txt) containing parameters determined using linear regression and .csv files containing allele-specific stutter ratios. The creation of each of these file types is described below.

2.2.1 Stutter regression files (.txt files)

The stutter regression text file describes the stutter ratio, *SR,* for each locus using the equation of a straight line where:

$$SR = \text{Slope} \times \text{Allele} + \text{Intercept}$$

This relationship describes an allele-specific stutter ratio based on allelic designation and is used to calculate the expected height of stutter and allelic peaks within STRmix™. The slope and intercept are locus-specific and may be determined using linear regression. An example of how this can be undertaken in MS Excel is provided below. Within the above equation, 'Allele' refers to the allelic designation of the allele being considered.

Where *SR* is observed to remain approximately constant regardless of allele size, the average observed *SR* across all alleles at the locus may be calculated and used to model expected *SR*. This approach may be suitable for stutter variants such as forward stutter, double back stutter, and half back stutter where *SR* typically does not vary according to allele size (with the exception of the D22S1045 locus where a linear relationship is often observed). A per locus average observed *SR* model can be implemented within STRmix™ by setting the slope to 0 and the intercept to the per locus average observed *SR* within the stutter regression text file.

Creating stutter regression files (.txt files)

A	B	C	D	E	F
Sample Name	Marker	Stutter	Stutter Height	Allele	Allele Height
Sample1	CSF1PO	9	30	10	979
Sample2	CSF1PO	9	30	10	840
Sample3	CSF1PO	9	30	10	833
Sample4	CSF1PO	7	31	8	1498
Sample5	CSF1PO	10	31	11	985
Sample6	CSF1PO	9	31	10	989
Sample7	CSF1PO	9	32	10	1282
Sample8	CSF1PO	12	32	13	535
Sample9	CSF1PO	12	34	13	675
Sample10	CSF1PO	11	35	12	776
Sample11	CSF1PO	11	35	12	788
Sample12	CSF1PO	9	37	10	910
Sample13	CSF1PO	9	38	10	1277
Sample14	CSF1PO	12	38	13	669
Sample15	CSF1PO	11	38	12	707

Analyse the data with stutter filters turned off. As stated above, it is recommended that analysis be undertaken using a reduced AT in order to detect more stutter data and better inform the models developed.

Retain labels for allelic peaks as well as for all stutter peaks for stutter variants that you wish to model. Remove labels from other artefact peaks (e.g. pull-up, drop-in etc.).

Import your data into MS Excel.

Omit all observations where the two alleles of a heterozygote are separated by one repeat unit (i.e. stutter affected heterozygotes).

Sort the remaining data by locus (marker) so that the parent allele and corresponding stutter peak(s) are in the same row, as shown.

Missing data	Where insufficient data has been observed for a particular stutter type

Even where a reduced AT has been used to analyse the data, stutter peaks may not be detected. This is particularly true for stutter variants that typically have low *SR*s (e.g. forward stutter, double back stutter etc.). The insertion or addition of data *in silico* is a common way that statisticians handle 'missing' data (in this context, stutter peaks that are present but below AT). Where there has been insufficient stutter data observed, if missing data is not accounted for, expected *SR*s may be overestimated. One way to account for missing data is to insert a stutter peak with some height below the AT used to analyse the data. A suggested approach is provided in the box to the right.

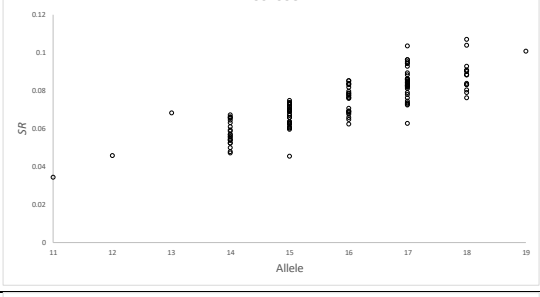
and missing data is of concern, an option is to take the following approach. Where parent peak height is >1000 rfu (3130 CE data) or >4000 rfu (3500 CE data) and the stutter variant being modelled has not been detected, insert a stutter peak for the relevant variant with height equal to half AT. Care should be taken to ensure this approach does not introduce underestimation or overestimation effects. This can be assessed after subsequent Model Maker analysis, by reviewing the correlation plots for any trends towards underestimation or overestimation for the relevant stutter types.

A	B	C	D	E	F	G
Sample Name	Marker	Stutter	Stutter Height	Allele	Allele Height	SR
Sample1	CSF1PO	9	30	10	979	0.030644
Sample2	CSF1PO	9	30	10	840	0.035714
Sample3	CSF1PO	9	30	10	833	0.036014
Sample4	CSF1PO	7	31	8	1498	0.020694
Sample5	CSF1PO	10	31	11	985	0.031472
Sample6	CSF1PO	9	31	10	989	0.031345
Sample7	CSF1PO	9	32	10	1282	0.024961
Sample8	CSF1PO	12	32	13	535	0.059813
Sample9	CSF1PO	12	34	13	675	0.05037
Sample10	CSF1PO	11	35	12	776	0.045103
Sample11	CSF1PO	11	35	12	788	0.044416
Sample12	CSF1PO	9	37	10	910	0.040659
Sample13	CSF1PO	9	38	10	1277	0.029757
Sample14	CSF1PO	12	38	13	669	0.056801
Sample15	CSF1PO	11	38	12	707	0.053748

Calculate:

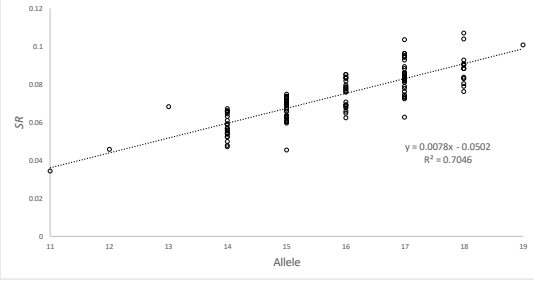
$$SR = \frac{O_{a\pm(i,j)}}{O_a}$$

Where $O_{a\pm(i,j)}$ is the observed peak height of the stutter peak and O_a is the observed peak height of the parent allele.



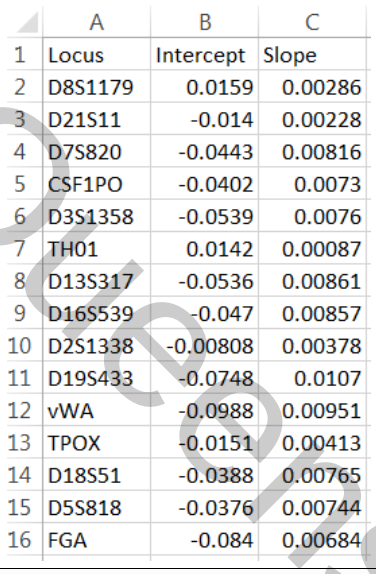
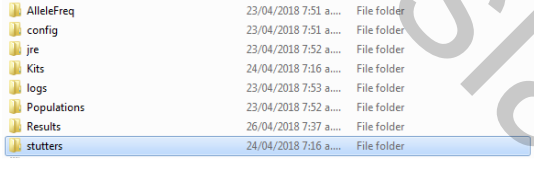
For each locus, plot *SR* versus allelic designation.

This provides a visualisation of the relationship between *SR* and allelic designation at each locus. Any apparent outliers should be investigated further.



Add a trend line to the plot. Right click the data series, then select **Add Trendline**. In the Format Trendline window, select **Linear** and ensure that **Display Equation on chart** and **Display R-squared value on chart** are ticked.

The slope and intercept can be read from the equation displayed. In the example shown, the slope is 0.0078 and the y-intercept is -0.0502.

	The R^2 statistic describes how well the trend line fits the data. In general, larger R^2 values are desirable.
	Repeat for all loci.
 A B C 1 Locus Intercept Slope 2 D8S1179 0.0159 0.00286 3 D21S11 -0.014 0.00228 4 D7S820 -0.0443 0.00816 5 CSF1PO -0.0402 0.0073 6 D3S1358 -0.0539 0.0076 7 TH01 0.0142 0.00087 8 D13S317 -0.0536 0.00861 9 D16S539 -0.047 0.00857 10 D2S1338 -0.00808 0.00378 11 D19S433 -0.0748 0.0107 12 vWA -0.0988 0.00951 13 TPOX -0.0151 0.00413 14 D18S51 -0.0388 0.00765 15 D5S818 -0.0376 0.00744 16 FGA -0.084 0.00684	Transcribe the values determined for intercept and slope for each locus into a new MS Excel file. The format of the stutter regression text file is: locus name, *y*-intercept, and slope. The locus names must be identical to those within the kit. Exemplar files are included with the STRmix™ installation for commonly-used kits and may be used as a template when creating your own files.
 AlleleFreq 23/04/2018 7:51 a.... File folder config 23/04/2018 7:51 a.... File folder jre 23/04/2018 7:52 a.... File folder Kits 24/04/2018 7:16 a.... File folder logs 23/04/2018 7:53 a.... File folder Populations 23/04/2018 7:52 a.... File folder Results 26/04/2018 7:37 a.... File folder stutters 24/04/2018 7:16 a.... File folder	Save as a Text (tab delimited) .txt file within the Stutters sub-folder in the STRmix™ V2.8 root directory.
	Repeat this process for each stutter variant you wish to model. Each stutter variant being modelled must have its own stutter regression text file saved within the Stutters sub-folder. As detailed above, a per locus average observed *SR* may be calculated and used to model stutter. This approach is not recommended for back stutter but may be suitable for other stutter variants where *SR* is not observed to vary with increasing allele size. To implement this approach within STRmix™, edit the regression parameters within the stutter .txt file so that the slope is equal to zero and the intercept is equal to the per locus average observed *SR* calculated.

2.2.2 Stutter Exceptions files (.csv files)

For some loci, regression against allelic designation as described above may not provide a good model to estimate expected stutter ratios. A better model may be developed by instead regressing *SR* against the longest uninterrupted stretch of

repeat units (*LUS*). This approach requires knowledge of sequence information for commonly observed alleles at the locus; such information may be obtained online, either from STRBase[1] or from the STRmix™ Technical and Scientific Support platform[2]. Note that access to the STRmix™ Technical and Scientific Support platform is restricted to those users/laboratories who have purchased STRmix™. Please email support@strmix.com for further information or to request access to the platform.

Within STRmix™, the *LUS* model for stutter can be implemented using a stutter exceptions file (a .csv file). This file provides a look-up table containing per allele *SR* values. Refer to the table below for further instructions.

Alternatively, for those loci where neither allelic designation nor *LUS* designation are observed to be good predictors of *SR*, the average observed *SR* for each allele at the locus may be calculated and transcribed into the stutter exceptions file. It is recommended that the average observed *SR* is only calculated for those alleles where sufficient stutter observations have been made (e.g. three or more observations).

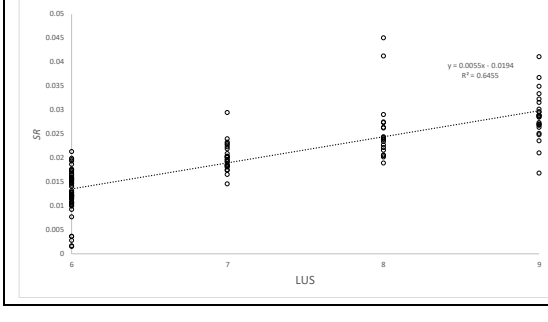
During interpretation, STRmix™ will first reference the stutter exceptions file. If no file has been saved or if a value is not available within the exceptions file for the allele of interest (indicated by a 0 within the exceptions file), STRmix™ will instead default to using the stutter regression text file to calculate expected *SR*.

Creating stutter exception files (.csv files)

A	B	C	D	E	F	G	H
Sample Name	Marker	Stutter	Stutter height	Allele	LUS	Allele height	SR
Sample1	CSF1PO	6	56	7	7	3491	0.016041
Sample2	CSF1PO	6	105	7	7	6533	0.016072
Sample3	CSF1PO	6	74	7	7	4398	0.016826
Sample4	CSF1PO	6	52	7	7	3022	0.017207
Sample5	CSF1PO	6	134	7	7	7737	0.017319
Sample6	CSF1PO	6	60	7	7	3442	0.017432
Sample7	CSF1PO	6	87	7	7	4838	0.017983
Sample8	CSF1PO	6	73	7	7	3868	0.018873
Sample9	CSF1PO	6	47	7	7	2489	0.018883
Sample10	CSF1PO	6	137	7	7	7173	0.019099
Sample11	CSF1PO	6	137	7	7	7133	0.019207
Sample12	CSF1PO	6	69	7	7	3565	0.019355
Sample13	CSF1PO	6	158	7	7	7745	0.0204
Sample14	CSF1PO	6	42	7	7	1916	0.021921
Sample15	CSF1PO	8	66	9	9	2183	0.030234
Sample16	CSF1PO	9	45	10	10	1485	0.030303
Sample17	CSF1PO	9	44	10	10	1435	0.030662

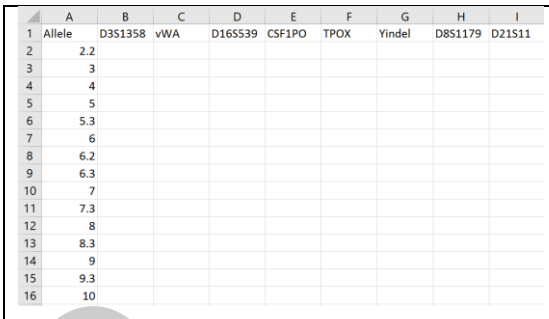
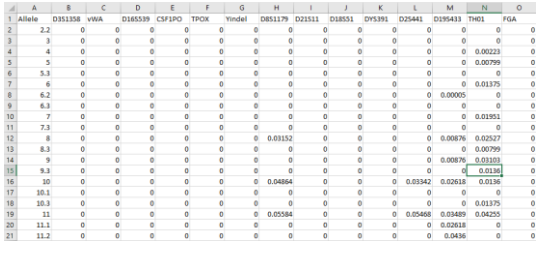
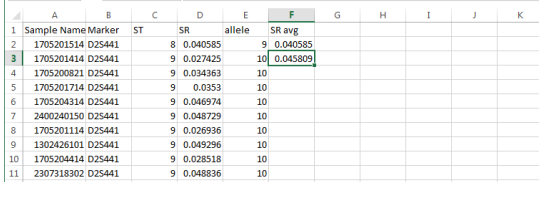
In MS Excel, add a column to your previous stutter workup. In this column, enter the *LUS* designation for each allele.

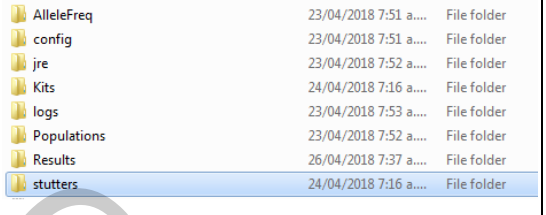
If multiple *LUS* values are available for an allele, one option is to take the average of these values.



Plot *SR* versus LUS designation. Insert a trend line as described above.

[1] https://strbase.nist.gov//str_fact.htm
[2] https://support.strmix.com (if necessary please sign in to https://support.strmix.com and navigate to: Solutions/Implementation and Validation/Implementation assistance/LUS look up table)

	Within MS Excel, create a new file. The first row is reserved for headers; the format is Allele, Locus 1, Locus 2…Locus n, where Locus n refers to the locus name. The locus names in the stutter exceptions file must be identical to those within the STRmix™ kit settings. For each allele, enter the allelic designation in a separate row in column 1. Exemplar files are included with the STRmix™ installation for commonly-used kits and may be used as a template when creating your own files.
Example calculation for TH01 9.3 allele: Trendline equation determined above is $SR = 0.0055x - 0.0194$ LUS for TH01 9.3 allele = 6 $SR = 0.0055 \times 6 - 0.0194 = 0.0136$ Transcribe this value into the stutter exceptions file in the cell for the TH01 9.3 allele: 	For those loci where *LUS* is observed to be a good predictor of *SR*, calculate the expected *SR* for each allele using the trend line equation determined above. The *LUS* designation of the allele being considered should be substituted for x. Transcribe the values calculated into the stutter exceptions file against the allelic designation of the allele being considered. Where no *LUS* value is available for a particular allele or if a negative value is calculated, enter a zero (0) against the allele. The file should contain those loci marked as Ignored in the STRmix™ kit file e.g. Y-indel and DYS391 within the GlobalFiler™ multiplex, however does not need to include those loci indicated in the kit file as Gender markers e.g. Amelogenin. The *SR* values for such loci should all be set to 0.
	For loci where *SR* is poorly described by both allele and *LUS* (e.g. SE33), empirical observations of *SR* values may be more suitable. For such loci, calculate the average observed *SR* for each allele. Transcribe

	these values into the stutter exceptions file against the corresponding allele.
	Save as a CSV (comma delimited) file (.csv) within the Stutters sub-folder in the STRmix™ root directory.

2.3 Saturation

The peaks in a DNA profile are measured using fluorescence. The amount of fluorescence is proportional to the quantity of DNA present. This fluorescence is captured by a camera. It is expected that as more DNA is added into a PCR the resulting peak intensity in an epg will increase. The camera can become saturated when too much fluorescence is detected. This means that the amount of DNA that is actually present can no longer be accurately measured. The saturation setting is the upper limit for peak height permitted within STRmix™. Saturation is largely instrument related and not kit or method dependent. The saturation threshold may be found by using a dilution series, or, in a subtler approach, by using the known relationship of parent allele to stutter. An example of how the saturation threshold can be determined is provided below.

Determining saturation

Sample File	Marker	O_{a-1}	O_{a-1} height	Allele	Height	SR
Sample1	D7S820	11	440	12	6550	0.0653
Sample2	D7S820	11	392	12	5850	0.0653
Sample3	D7S820	11	144	12	2115	0.0653
Sample4	D7S820	11	275	12	4000	0.0653
Sample5	D7S820	11	238	12	3339	0.0653
Sample6	D7S820	11	233	12	3562	0.0653
Sample7	D7S820	11	421	12	6399	0.0653
Sample8	D7S820	11	371	12	5461	0.0653
Sample9	D7S820	11	41	12	294	0.0653
Sample10	D7S820	11	250	12	3559	0.0653

Analyse a range of profiles of varying DNA template amount including over-amplified profiles. Analyse the data with stutter filters turned off. Retain labels for all allelic and back stutter ($O_{a-1,0}$) peaks.

Remove stutter affected heterozygotes. Sort the data by locus (marker) so that the parent allele and corresponding back stutter are in the same row, as shown.

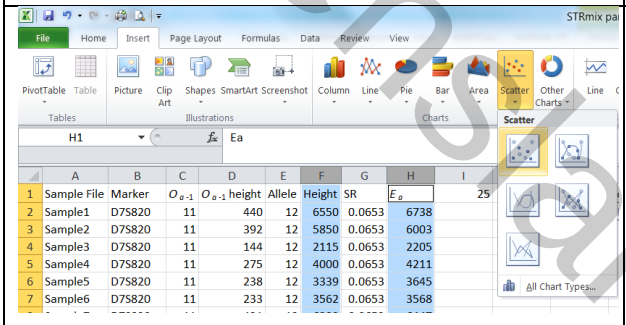
Filter the data so that only loci with simple repeat structures are listed, where the Allele linear regression is a good predictor of SR. Simple loci include CSF1PO, D13S317, D16S539, D18S51 (excluding the x.2 microvariant alleles), D10S1248, D5S818, D7S820, Penta D, Penta E, and TPOX.

Sample File	Marker	O_{a-1}	O_{a-1} height	Allele	Height	SR	E_a
Sample1	D7S820	11	440	12	6550	0.0653	6738
Sample2	D7S820	11	392	12	5850	0.0653	6003
Sample3	D7S820	11	144	12	2115	0.0653	2205
Sample4	D7S820	11	275	12	4000	0.0653	4211
Sample5	D7S820	11	238	12	3339	0.0653	3645
Sample6	D7S820	11	233	12	3562	0.0653	3568
Sample7	D7S820	11	421	12	6399	0.0653	6447
Sample8	D7S820	11	371	12	5461	0.0653	5681
Sample9	D7S820	11	41	12	294	0.0653	628
Sample10	D7S820	11	250	12	3559	0.0653	3828

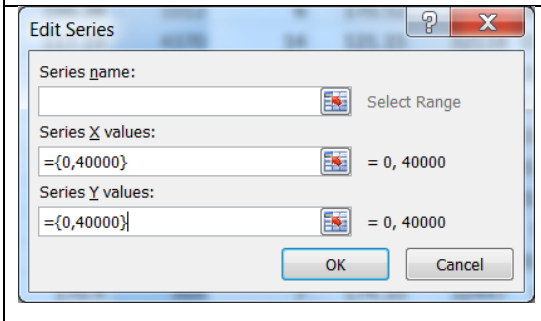
Calculate:

$$E_a = \frac{O_{a-1,0}}{SR_{a-1,0}}$$

Where E_a is the expected allele peak height, $O_{a-1,0}$ is the observed back stutter peak height, and $SR_{a-1,0}$ is the expected back stutter ratio for allele a calculated using the linear equation determined in Section 2 above (SR = Slope × Allele + Intercept). The slope and intercept values used will be those saved in the stutter regression text file prepared in Section 2.2.1.

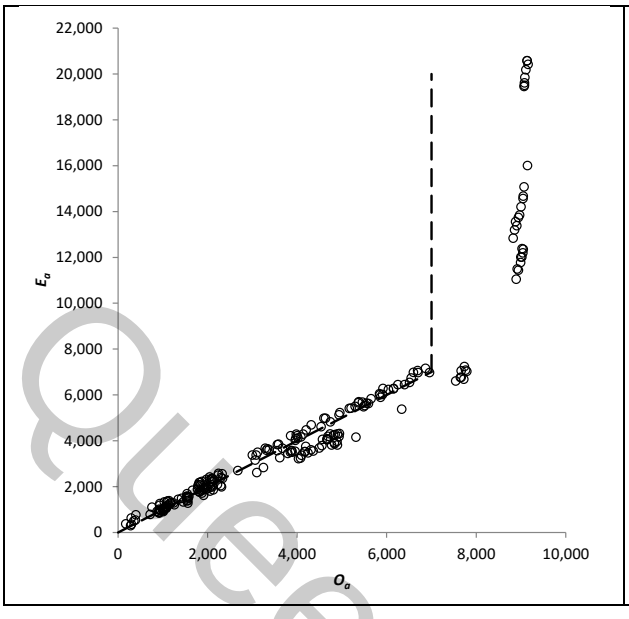


Plot E_a, the expected allele peak height calculated above, against observed allele peak height (O_a)



Also plot a line for $x = y$. One way to do this is to add a new series. Right click on the plot and choose Select Data > Add and then enter appropriate values.

The values shown in the screenshot to the left will plot data points at (0,0) and (40000,40000). Right click on one of these data points and select **Add Trendline**.



The points should start to deviate from the $x = y$ line at the point of saturation as demonstrated in the figure to the left. Select a saturation value that sits just below the saturation point observed. In the example shown, a saturation threshold of 7,000 rfu has been selected and is indicated in the plot using a vertical dashed line.

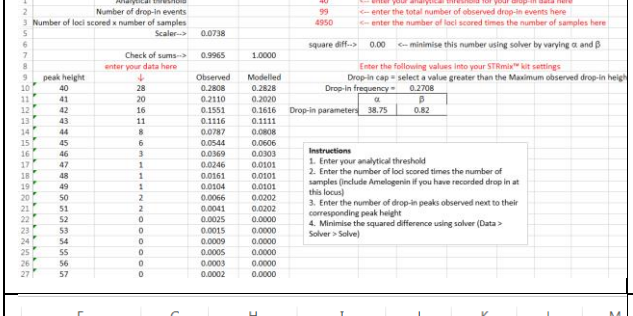
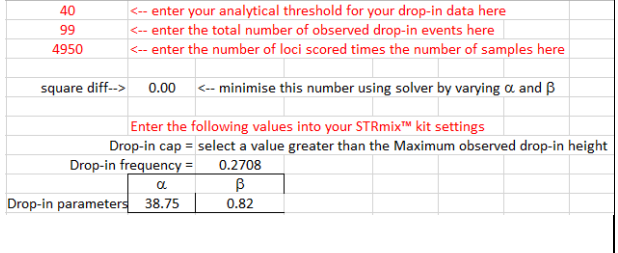
Common saturation values observed are 7,000 rfu for a 3130 instrument and 30,000 rfu for a 3500 instrument.

2.4 Drop-in

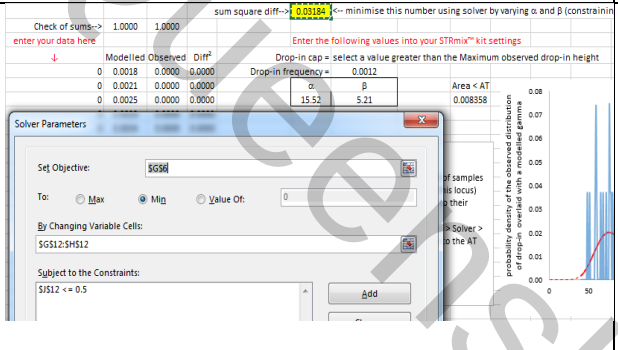
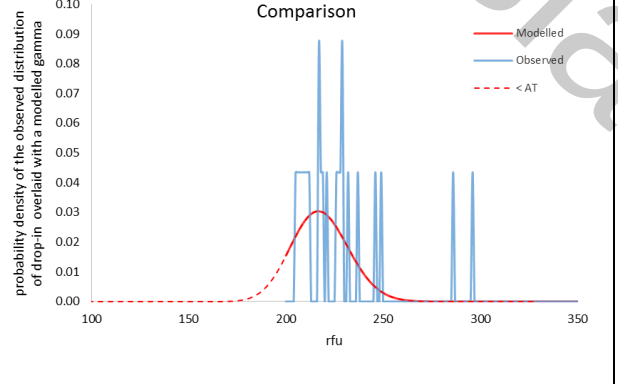
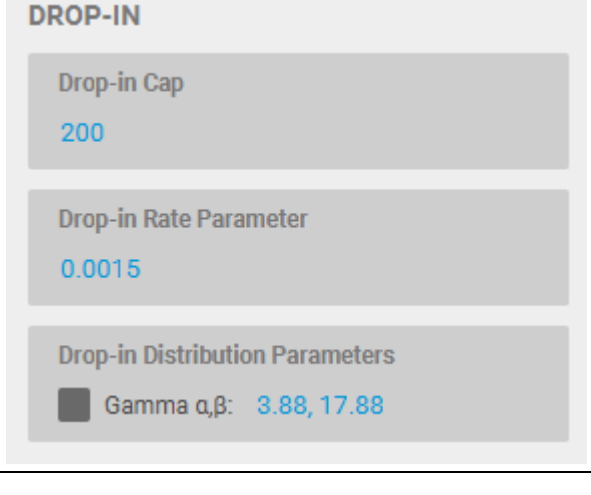
Drop-in is defined as the observation of non-reproducible, unexplained peaks within a DNA profile. Typically, this manifests as one or two low-level peaks. There are four parameters used for the modelling of drop-in within STRmix™. These are:

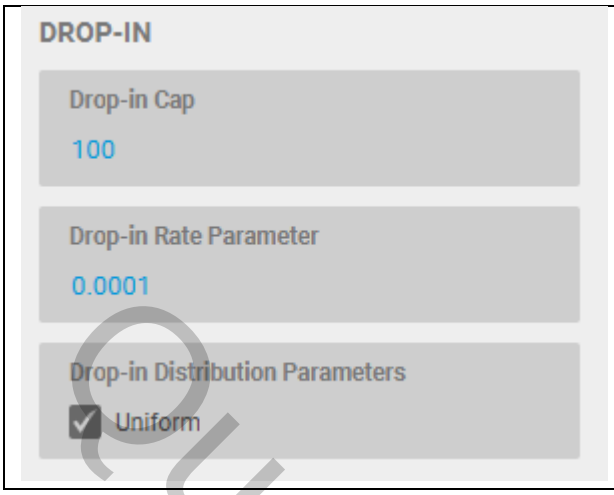
1. The analytical/detection threshold,
2. A cap on the maximum allowed height for a proposed drop-in peak,
3. The drop-in rate parameter, and
4. α, β: two parameters that describe a gamma distribution.

Drop-in rates for a laboratory platform (multiplex and instrument combination) should be monitored. This is done by recording counts and corresponding heights of drop-in peaks observed in negative controls along with a count of the total number of negative controls examined. Within STRmix™, drop-in can be modelled using a gamma distribution. The determination of the parameters for a gamma model is discussed below. If limited observations of drop-in have been made or if the dataset is limited in size, it may not be possible to reliably develop a gamma model. In these instances, it is recommended that a uniform distribution based on drop-in frequency is instead used.

Drop-in worksheet	
Gamma model	The drop-in worksheet can be accessed via the Implementation Assistance tab on the STRmix™ Technical and Scientific Support platform[3].
	Within the workbook enter the counts of observed drop-in peaks in column B. Ensure that the peak height range in column A extends beyond the highest drop-in peak height you have observed. It is recommended that negative controls are analysed using a reduced AT to better detect instances of drop-in.
	Complete the information in column F. Cell F1: Enter the AT used to analyse the negative controls. Cell F2: Enter the total number of drop-in events observed. Cell F3: Enter the number of loci in your multiplex kit multiplied by the total number of negative controls examined. If gender markers or non-autosomal loci (e.g. Amelogenin, Y-indel, DYS391) were examined for drop-in then these

[3] https://support.strmix.com (if necessary please navigate to: Solutions/Implementation and Validation/Implementation assistance/Drop-in calculator)

	loci should be included when considering the number of loci scored.
	Minimise the squared differences using the Excel Add-in tool, Solver (Data > Solver > Solve) by varying α and β subject to the constraint that Area < AT (cell J12) must be ≤ 0.5. If an error message is encountered, set α equal to the AT and β equal to 1 and run Solver again.
	A chart plots the observed versus modelled drop-in distribution for the dataset. This can be used to check the modelling performance.
	Enter the drop-in parameters into the Profiling Kits screen of STRmix™ for the appropriate kit. The drop-in cap is the maximum permitted drop-in peak height. This should be set to some value that is larger than the maximum height observed for a drop-in peak during validation. The drop-in frequency and distribution parameters are determined using the worksheet as described above.

	If using a uniform model, tick the box within the Drop-In Distribution Parameters field. The screen will update to indicate that a Uniform model has been selected (see screenshot to left). Follow the instructions above to determine the drop-in frequency and drop-in cap.

2.5 Peak height variance and LSAE variance parameters

Traditionally, a 'bottom up' approach to setting variance constants is taken, with these parameters derived by modelling the observed variability in a range of single-source profiles. The appropriateness of these values is then assessed using a series of diagnostics and their performance tested, in combination with the STRmix™ parameters already described, by conducting specificity and sensitivity testing as outlined in section 3.5 Section D - Sensitivity and specificity.

The bottom up approach to setting variance parameters involves measuring the variability in a number of single-source profiles of known origin using the Model Maker function within STRmix™. The profiles used should encompass the broad range of profile qualities likely to be encountered in casework, from low-level partial profiles to full profiles approaching the CE camera's saturation threshold. Note that profiles with 10 or fewer peaks and profiles containing peaks above the saturation threshold will be ignored by Model Maker. In STRmix™ V2.8, Model Maker is used to determine the peak height variance prior distributions for allelic and stutter peaks, along with the LSAE variance prior distribution. A prior distribution is determined for each stutter variant being modelled. Refer to the STRmix™ V2.8 User's manual for further details on how Model Maker works.

Drop-in modelling needs to be turned off when performing STRmix™ V2.8 Model Maker analysis (unless otherwise specified in the Release and Testing reports). This can be done by setting all drop-in modelling parameters to 0 within a kit file prior to its selection in Model Maker set up. Release and Testing reports are available on the STRmix™ Technical and Scientific Support platform[4].

[4] https://support.strmix.com (if necessary please navigate to: Solutions/Implementation and Validation/Release and Testing Reports – STRmix™)

Setting up kits and Model Maker

Acceptable single-source profiles input file formats:

Sample Name	Marker	Allele1… Allele *n*	Size1… Size *n*	Height1… Height *n*

Sample Name	Marker	Allele1	Size1	Height1	Allele2…

Single-source profiles input file:

For each kit and instrument combination, prepare a set of at least 90 single-source profiles. These profiles should originate from known donors and span the broad range of profile qualities observed in casework samples. One approach is to prepare a dilution series and supplement this with profiles exhibiting a higher degree of stochastic peak height variation e.g. artificially degraded samples. Alternatively, non-probative casework profiles may be used where the donor can be reasonably inferred (e.g. complainant's profile on an intimate swab, fingernails, clothing etc.).

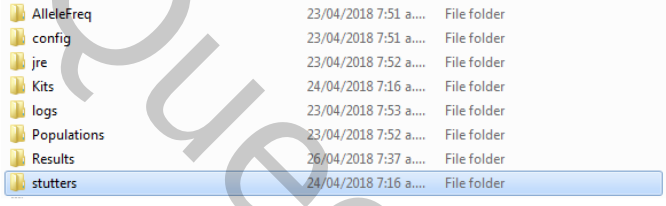
Efforts should be made to avoid a dataset that is too 'pristine' and homogenous with respect to peak height, bearing in mind that any profiles containing ≤ 10 peaks and any profiles with peaks above the saturation threshold will be ignored.

Analyse the profiles within your analysis software with stutter filters turned off. Retain labels for all allelic peaks along with peaks for all stutter variants being modelled. It is recommended that a reduced AT is used to detect more of the DNA data within the profiles. Labels should also be retained for apparent drop-in peaks. Remove labels from all other peaks (pull-up, spikes etc.).

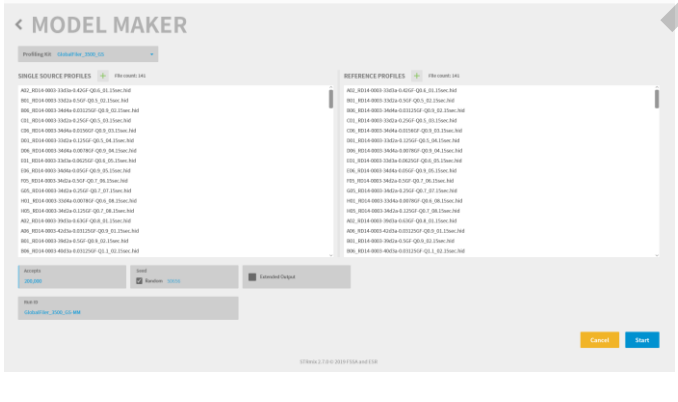
Create a STRmix™ .txt evidence input file by exporting sample name/sample file, marker, allele, size, and height. Examples of acceptable file formats are shown in the panel to the left. Refer to the STRmix™ V2.8 Operation Manual for further instructions.

Peaks that cannot be explained by the reference profile, or as stutter peaks will be flagged by Model Maker at the start of the analysis and will require resolution before the analysis will run.

<table>
<tr>
<td>Reference profiles input file format:
<table><tr><td>SampleFile</td><td>Marker</td><td>Allele1</td><td>Allele2</td></tr></table></td>
<td>Reference profiles input file:

Prepare a second file containing the known genotypes of each sample from the single-source profiles dataset. The sample names and order must match. We recommend copying and pasting the SampleFile and Marker columns from the single-source profiles input file and then completing the known genotype for each donor.</td>
</tr>
<tr>
<td> 
</td>
<td>Prior to Model Maker interpretation in STRmix™, a kit must be created containing all of the settings determined in the above sections (stutter files, analytical thresholds, drop-in, saturation etc.). If the Model Maker dataset has been analysed using a reduced AT, ensure that the Detection Threshold settings are changed to correspond with the AT used.

Stutter files should previously have been saved in the Stutters sub-folder of the STRmix™ root directory.

Refer to the STRmix™ V2.8 Operation Manual for instructions on how to set up kits within STRmix™.</td>
</tr>
<tr>
<td> 
</td>
<td>From the main menu, Select Model Maker to open the Model Maker window.

Select the appropriate kit from the Profiling Kit drop down menu.

Drag and drop the single-source profiles and reference profiles input files into the appropriate windows.

If desired, change the Run ID to rename your Model Maker run.

The recommended number of Accepts is 200,000.

Select Start to begin the Model Maker interpretation.

The analysis may take a few hours depending on the specifications of the computer, the number of profiles, and the number of stutter variants being modelled.</td>
</tr>
</table>

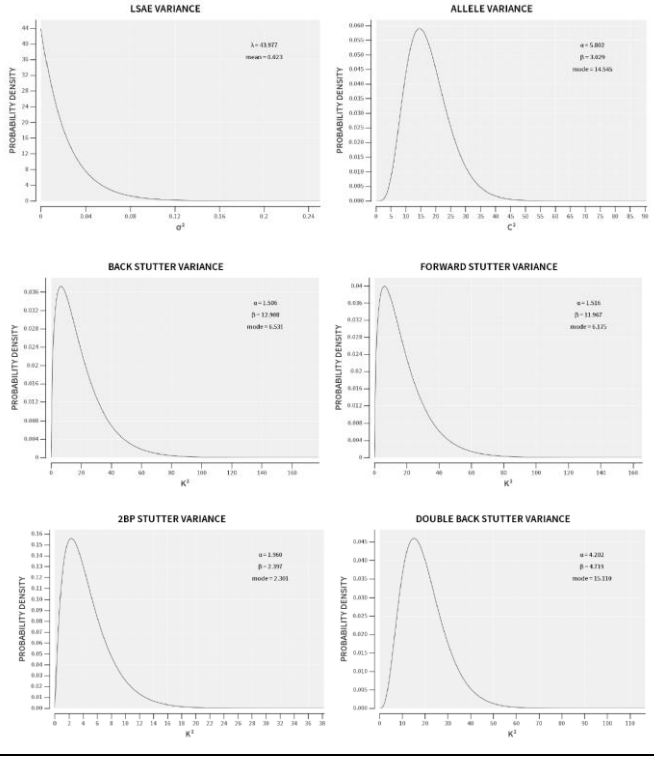
PROBABILITY DISTRIBUTION

	α	β	MODE
Allele Variance c^2	5.802	3.029	14.545
Back Stutter Variance k^2	1.506	12.908	6.531
Forward Stutter Variance k^2	1.516	11.967	6.175
2bp Stutter Variance k^2	1.960	2.397	2.301
Double Back Stutter Variance k^2	4.202	4.719	15.110
LSAE Variance	0.023		

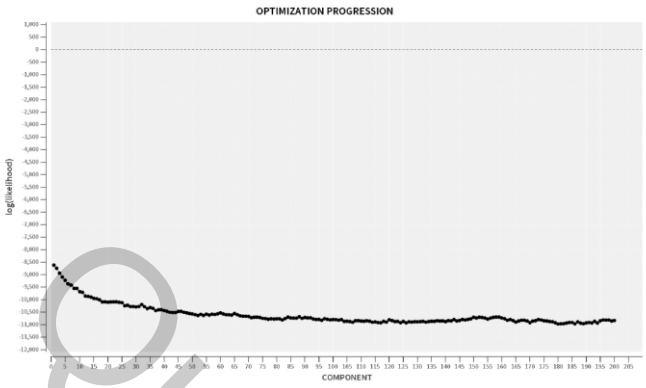
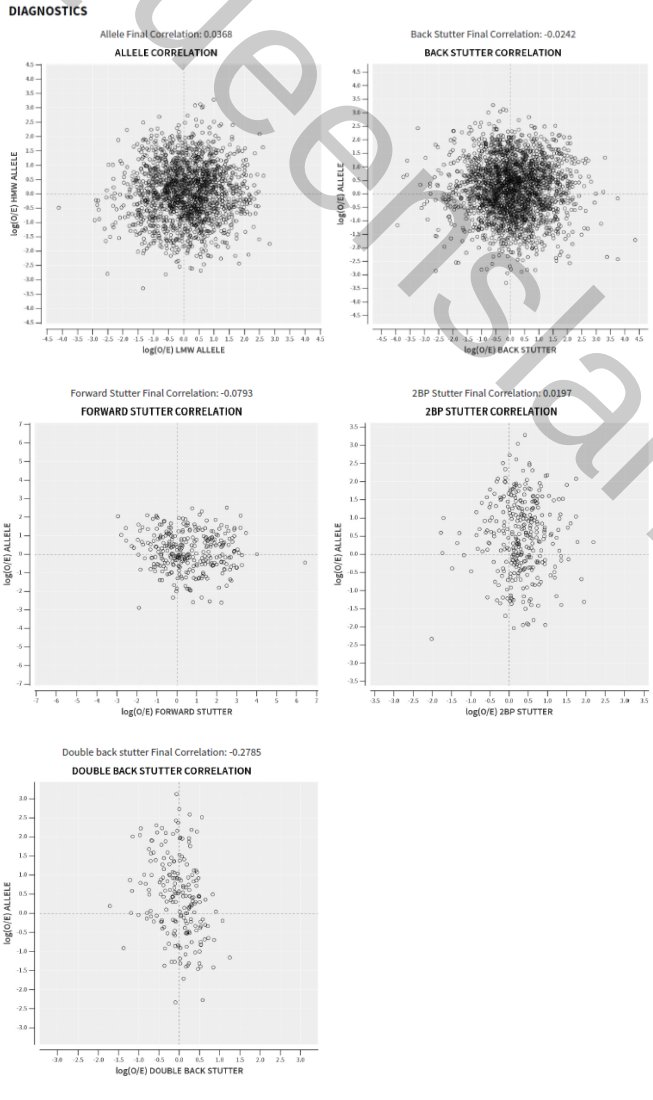
The Model Maker results are saved in a folder within the Results directory.

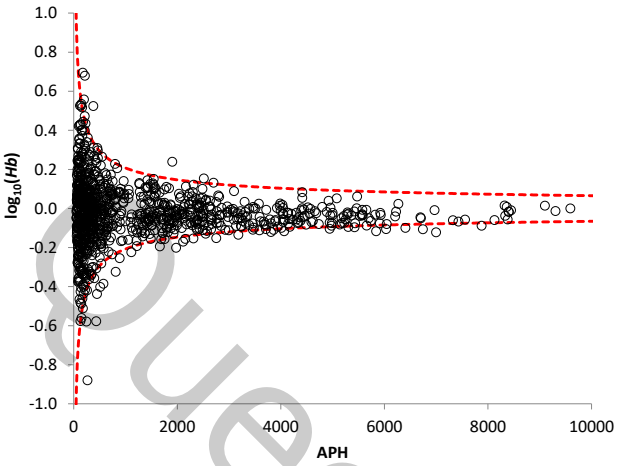
Parameters describing the peak height variance prior distributions for allelic and stutter peaks are provided in the Model Maker PDF report. Within STRmix™, peak height variance is modelled using a gamma distribution. Individual parameters will be listed for each stutter variant being modelled. Also provided is the mean of the LSAE variance prior distribution (modelled using an exponential distribution).

Within STRmix™ > Administration > Profiling Kits, select the relevant kit, then select the **Import** tab. Drag and drop the Model Maker results folder to the Model Maker Result field, then click the **Import** button to copy the relevant parameters to the kit settings. Alternatively, click **Browse**, navigate to the Model Maker results folder, select the config.xml file, then click **Import**. Note that the kit name selected must match the kit name used when running Model Maker. Alternatively, transcribe the variance parameters to the relevant fields within the **General** and **Stutters** tabs.



The allele and stutter peak height variance prior distributions are plotted within the Model Maker report. A plot is also provided for the LSAE variance prior distribution.

Log(likelihood) values 	The log(likelihood) of the dataset at the current component-wise MCMC cycle (in increments of 1000) is also provided in the Model Maker report. This is shown to ensure that the analysis has converged, which will be indicated by a plateauing of log(likelihood) values. The plot to the left shows the log(likelihood) for one dataset over the course of the Model Maker run. In this dataset, the results plateau after approximately 80 components. Note that the data may plateau either up or down.
	Correlation plots are also included within the Model Maker report; it is recommended that these plots be reviewed for any correlation. Ideally, no correlation will be observed. These plots may also indicate any outliers requiring further review. The correlation plots for some stutter variants may have limited data points depending on the number of stutter peaks of that type observed within the Model Maker dataset. Refer to the STRmix™ V2.8 User's Manual for further information regarding the correlation plots provided within the Model Maker report.

Bounds on log(*Hb*) versus APH  	A check of the output should be made. One approach is to plot the log of heterozygote balance (*Hb*) versus average peak height (APH) for each heterozygous locus within the Model Maker dataset, where: $\log(Hb) = \log\left(\frac{O_{HMW}}{O_{LMW}}\right)$ and $APH = \frac{O_{HMW} + O_{LMW}}{2}$ Within the above formulae, O_{HMW} refers to the observed height of the high molecular weight allele while O_{LMW} refers to the observed height of the low molecular weight allele. Plot these data along with the lines $\log(Hb) = \pm\sqrt{2} \times 1.96\sqrt{\frac{c^2}{APH}}$, where c^2 is a chosen percentile of the prior gamma distribution for allelic peak height variance (e.g. 75th percentile). The plotted lines should capture over 95% of the data. An example plot is shown in the figure to the left. A spreadsheet is available from the STRmix™ Technical and Scientific Support platform[5] to assist with this process. Follow the instructions included within the spreadsheet.

2.6 Populations and allele frequency files

Within STRmix™, population files contain population-specific information such as the default theta/F_{ST} value, priors for the unified *LR,* and population proportions if calculating a stratified *LR*. Introduced in V2.8, the minimum resampled count to use in the HPD *LR* can also be controlled from within the population file. The population file also directs STRmix™ to the relevant allele frequency file. Allele frequency files for a number of commonly-used populations (e.g. FBI extended and NIST populations) are included within the STRmix™ installation. Allele frequency files for additional populations can be readily created from the MS Excel tables commonly provided with modern population data publications. The locus names contained within the allele frequency file must match those within the STRmix™ kit settings. Allele frequency files may contain additional loci that are not present within the

[5] https://support.strmix.com (if necessary please navigate to: Solutions/Implementation and Validation/Implementation assistance/MM data to APH and Hb spreadsheet)

multiplex kit(s) used within the laboratory. Refer to the STRmix™ V2.8 Operation Manual for details on how to create a new allele frequency file. The STRmix™ V2.8 Operation Manual also contains instructions on how to create and edit populations within STRmix™.

3 INTERNAL VALIDATION FOLLOWING THE SWGDAM GUIDELINES

3.1 Preamble

STRmix™ is written in two halves. The first half uses Markov chain Monte Carlo (MCMC) processes to generate genotype weights while the second half uses those weights to assign a likelihood ratio. Each of these halves has undergone extensive *developmental validation*. This involved, in part, 'by hand' confirmation of the calculations used within the software. The results of the developmental validation are available on request, contact support@strmix.com for details if required. In addition, a summary of the developmental validation is discussed in Taylor et al. [7]. The developmental validation of STRmix™ following the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems [8] is described by Bright et al. [9]. Finally, a list of all papers describing the theory behind different aspects of STRmix™ is provided in Appendix 1 of this document.

Internal validation describes the activities a laboratory undertakes in-house before the implementation of the software into routine casework.

A *performance check* is a subset of the tests undertaken as part of internal validation to check the software is performing as expected. A performance check may be undertaken in lieu of internal validation between different versions of the software if the changes are not substantial. Tips for undertaking performance checks within STRmix™ are given in Appendix 2.

This document follows the internal validation section of the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems [8]. It describes a series of well-designed experiments that can serve to address multiple recommendations. A summary of the recommendations and in what section they are discussed is provided in Appendix 3.

Your laboratory system should be tested using representative data generated in-house with the same protocols and configuration of amplification kit, detection instrumentation, CE data analysis software as used for casework (recommendation 4.1). This includes any variable DNA typing conditions used within your laboratory (e.g. any variations in the amplification and/or electrophoresis parameters and/or CE data analysis methods used to increase the detection of alleles and/or decrease the detection of artefacts, recommendation 4.1.3). In addition, it should include specimens with known contributors as well as case-type specimens that may include unknown contributors (recommendation 4.1.1).

All internal validation experiments should be undertaken using parameters determined by your laboratory. The profiles used within the validation should differ from those used to determine laboratory specific parameters (e.g. the profiles used to run Model Maker) (recommendation 4.1.12).

Some guidance for the internal validation of PCR replicates, multi-kits, the variable number of contributors (varNOC) function, the M_x priors feature, mix to mix matching, and the top-down approach is given in Section 4. These features are not covered under any of the SWGDAM recommendations.

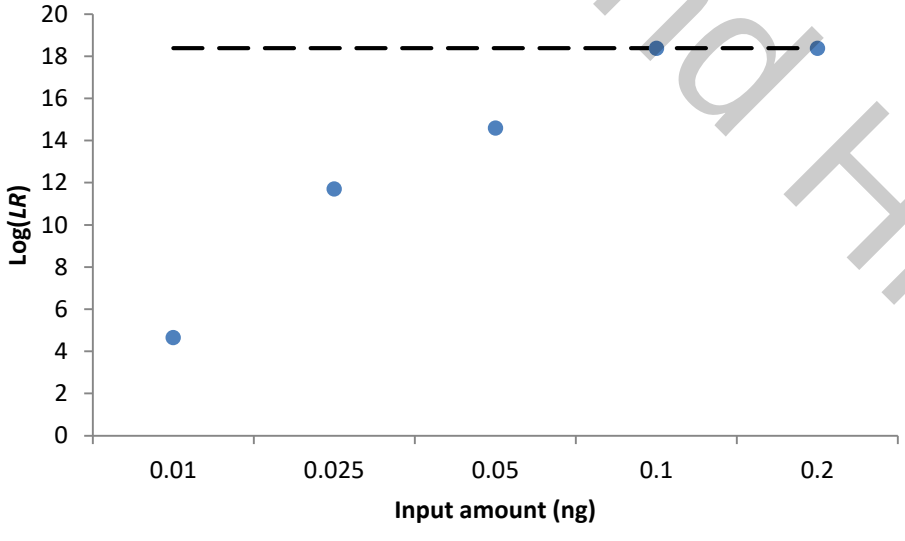
3.2 Section A – Single-source profiles

This section covers the following recommendation:

> 4.1.5. Single-source specimens
>
> 4.2.1.2. For single-source specimens with high quality results, genotypes derived from non-probabilistic analyses of profiles above the stochastic threshold should be in complete concordance with the results of probabilistic methods.

This section is intended to check whether the weights assigned to different genotype combinations are appropriate. The weights are the primary output of a STRmix™ deconvolution and should be intuitively correct, with those genotypes that best explain the observed profile being assigned relatively high weight. In contrast, genotypes that offer a poor explanation of the recovered profile should be assigned relatively little weight (or no weight at all).

Construct a dilution series of a single-source profile where the peak heights range from above the level where dropout is observed (i.e. above your laboratory's stochastic threshold) to below. Analyse the profiling data using your laboratory's analytical threshold. Peak labels should be retained for allelic peaks as well as for all stutter variants being modelled. Interpret the profiles in STRmix™, assigning an *LR* for the known contributor, and plot the resulting *LR*s. An example plot is provided below. The dashed horizontal line represents the log(*LR*) produced for a full, unambiguous, single-source profile from the known donor; this should equal the inverse of the random match probability of the donor's profile when calculated under the same conditions (i.e. using equations 1, 2, and 3 below).



The *LR* should trend from the single-source *LR* calculated for a full profile towards *LR* = 1 (log(*LR*) = 0) as DNA template decreases. The weights for genotypes considering dropout ([*a*, -1]) should increase as template decreases. The DNA amounts reported in the STRmix™ output (template) should also decline steadily in line with peak height.

For those profiles with peaks above your laboratory's previously assigned stochastic threshold, a manual interpretation of the profile should be in complete concordance with interpretation using probabilistic software and can simply be recorded as such.

In an extension of this experiment, the *LR* may be replicated by hand for a single-source profile where a single genotype has been assigned a weight is 1 (i.e. 100%) at each locus. The likelihood ratio calculation can be replicated by hand using the Balding and Nichols [10] formulae (recommendation 4.2 of NRCII and equations 4.10a and 4.10b therein). For single-source profiles:

$$\frac{2\left[\theta+(1-\theta)\,p_i\right]\left[\theta+(1-\theta)\,p_j\right]}{(1+\theta)(1+2\theta)}$$ for heterozygous loci Equation (1)

$$\frac{\left[2\theta+(1-\theta)\,p_i\right]\left[3\theta+(1-\theta)\,p_i\right]}{(1+\theta)(1+2\theta)}$$ for homozygous loci Equation (2)

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

$$\frac{x_i+\frac{1}{k+1}}{N_a+1}\quad\frac{x_i+\frac{1}{k+1}}{N_a+1}$$ Equation (3)

Where x_i is the number of observations of allele i within the allele frequency database, N_a is the total number of alleles typed at the locus under consideration within the database, and k is the number of different alleles observed at the locus under consideration within the database.

Setting θ to zero returns the product rule where:

$2p_ip_j$ for heterozygous loci Equation (4)

p_i^2 for homozygous loci Equation (5)

Using STRmix™, interpret a single-source profile where dropout and drop-in are <u>not</u> a consideration and assign an *LR* under the following propositions:

H_p: The DNA originates from the POI
H_d: The DNA originates from an unknown, unrelated individual

The *LR* assigned at each locus should be the same as that calculated by hand using equations (1), (2), and (3).

Each locus compared is an individual check of the maths. MS Excel spreadsheets implementing the above formulae are available from the STRmix™ Technical and Scientific Support platform[6].

3.3 Section B - Saturated data

This section covers the following standard:

4.1.4. Allelic peak height, to include off-scale peaks

[6] https://support.strmix.com (if necessary please navigate to: Solutions/Implementation and Validation/Implementation assistance/Single Source LR Calculator for STRmix v2.8)

Observed peaks within an electropherogram may be off-scale if they are above the saturation threshold calculated for a CE instrument (typically 7000 rfu for an Applied Biosystems 3130 instrument or 30,000 rfu for an Applied Biosystems 3500 instrument). This means that the allelic peak height is not accurately captured and therefore the observed stutter peak heights will be larger than their expected values. In this case, the relationship between stutter and allelic peak heights (i.e. stutter ratio) is unreliable. For this reason, where an allelic peak presents with height above the saturation threshold, expected stutter peak height is instead calculated within STRmix™ using *expected* allelic peak height rather than observed allelic peak height. It is recommended that grossly saturated profiles are not interpreted using STRmix™.

To review the impact of saturated data on profile interpretation, prepare a number of single-source profiles with peak heights exceeding the saturation threshold determined in Section 2.3 above. Interpret these profiles using STRmix™ and review the genotype weights. Additionally, review the k^2 diagnostic for back stutter. For very saturated profiles this diagnostic will generally sit out in the right hand tail of the prior distribution, indicating that larger than typical variation between observed and expected stutter peak heights was seen.

3.4 Section C - Mixture weights

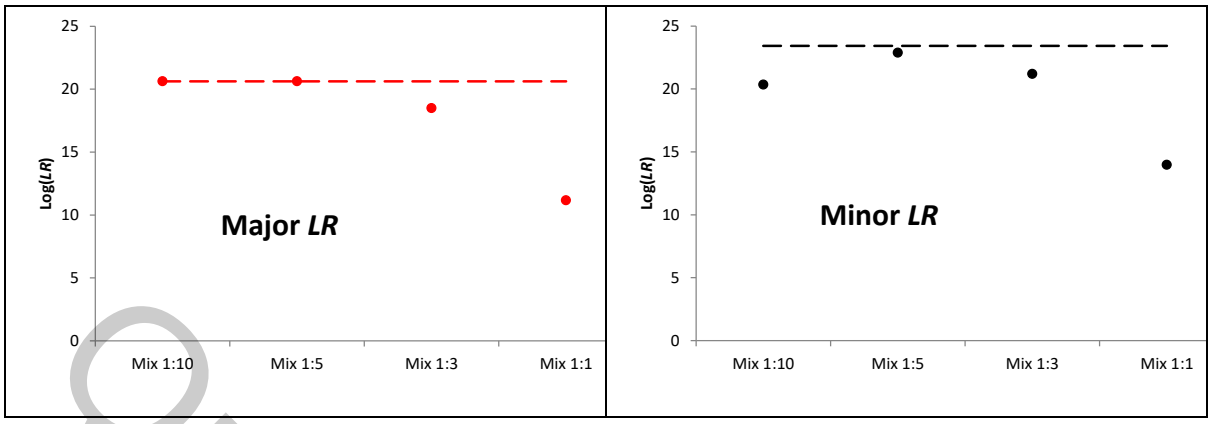
This section covers the following recommendation:

> 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.

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

Construct a two-person mixture series in the following ratios: 10:1, 5:1, 3:1, 2:1, and 1:1, aiming for a total input amount of DNA that aligns with your laboratory protocol. Analyse the DNA profiling data using your laboratory's casework AT, retaining labels for all stutter variants being modelled. Interpret the profiles in STRmix™, assign *LR*s for the major and minor contributors, and plot the resulting *LR*s.

We expect the *LR* to decrease for those mixtures where the major and minor contributor are unable to be resolved. The mixture proportions in the STRmix™ output should also change as the mixture ratio varies. Example plots are provided below. Within these plots, the log(*LR*) expected for a full, unambiguous, single-source profile from each donor has been plotted using a dashed line; the *LR*s observed should never be above this line.



The aim of the test is to ensure that the weights are intuitively correct. The weights within STRmix™ are the primary output of an interpretation. Non-intuitive weights are an indication of poor biological modelling. If STRmix™ behaves as described, it is performing as expected.

3.5 Section D - Sensitivity and specificity

This section covers the following recommendations:

- 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

With respect to interpretation methods, sensitivity is defined as the ability of the software to reliably resolve the DNA profiles of true contributors within a mixed DNA profile for a range of starting DNA templates. The log(*LR*) for 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, the use of conditioning profiles (for example, the complainant's profile when interpreting profiles recovered from intimate samples), the use of PCR replicates, and decreasing number of contributors (i.e. decreasing profile complexity). Specificity is defined as the ability of the software to reliably exclude 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 may be examined by interpreting a number of mixed DNA profiles within STRmix™ and assigning *LR*s for known contributors and non-contributors. Below we

provide a suggested approach covering mixtures of up to four contributors however this could be expanded upon if you wish to interpret more complex profiles (e.g. five-person mixtures). The profiles examined should span the range of DNA quantities and mixture proportions likely to be encountered in casework. Ensure that the contributors include homozygous and heterozygous genotypes and that there is varying amounts of allele sharing across the different loci (recommendation 4.1.6.5). Allele and/or locus dropout is expected to occur within those profiles that contain one or more contributors with low template amount (recommendation 4.1.7.1).

Four-person mixtures	Three-person mixtures	Two-person mixtures	DNA amount of smallest contributor (pg)
4:3:2:1 and 10:5:2:1	10:5:1 and 3:2:1	20:1, 10:1, 5:1, and 3:1	100
4:3:2:1 and 10:5:2:1	10:5:1 and 3:2:1	20:1, 10:1, 5:1, and 3:1	50
4:3:2:1 and 10:5:2:1	10:5:1 and 3:2:1	20:1, 10:1, 5:1, and 3:1	25
4:3:2:1 and 10:5:2:1	10:5:1 and 3:2:1	20:1, 10:1, 5:1, and 3:1	12.5
4:3:2:1 and 10:5:2:1	10:5:1 and 3:2:1	20:1, 10:1, 5:1, and 3:1	6.25

Four-person mixtures	Three-person mixtures	Two-person mixtures	DNA amount per contributor (pg)
1:1:1:1	1:1:1	1:1	400
1:1:1:1	1:1:1	1:1	200
1:1:1:1	1:1:1	1:1	100
1:1:1:1	1:1:1	1:1	50
1:1:1:1	1:1:1	1:1	25
1:1:1:1	1:1:1	1:1	12.5

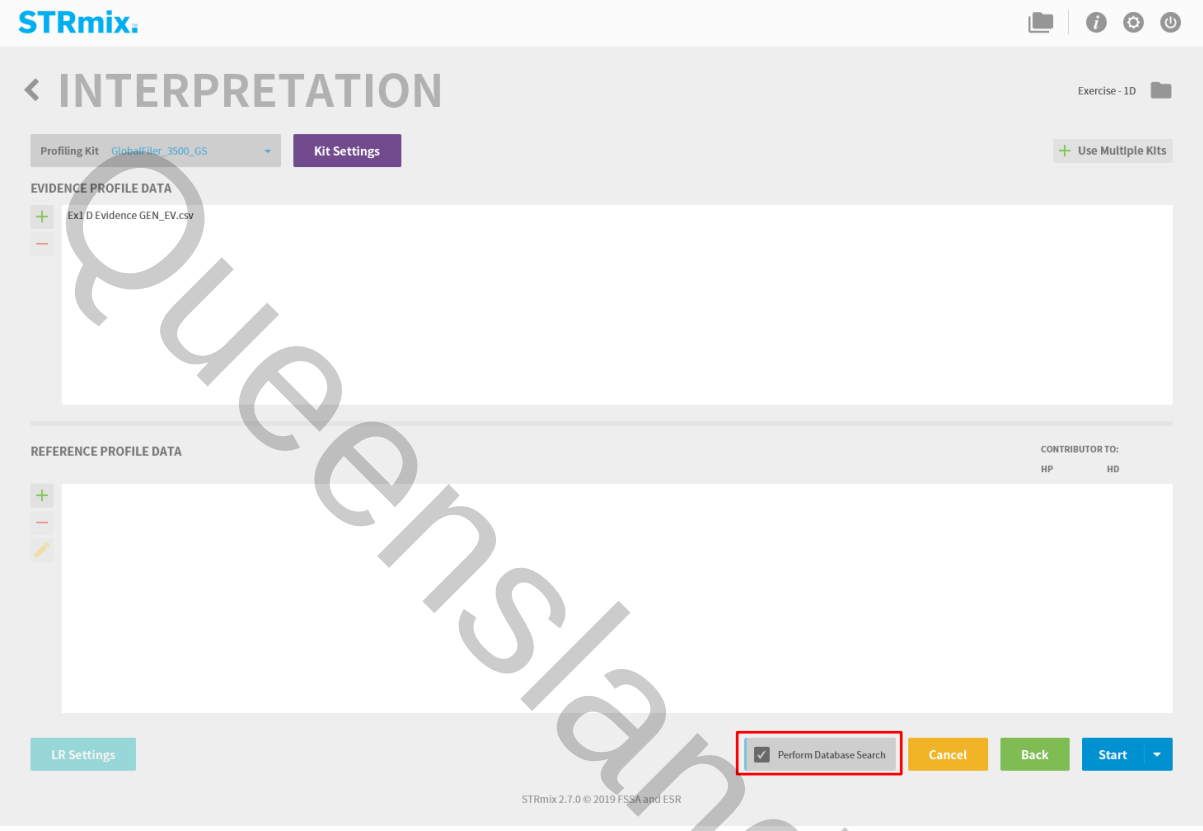
We recommend extending the above experimental design to test larger peak height ratio differences encountered in casework. One suggested approach is given below. For these large ratios, the M_x priors function can be useful. Re-interpret any counterintuitive results using this function. Refer to Section 4.4 for guidance on how to validate M_x priors.

Four-person mixtures	Three-person mixtures	Two-person mixtures	DNA amount of smallest contributor (pg)
100:100:100:6	100:100:4	100:2	25
100:100:100:6	100:100:4	100:2	12.5
100:100:100:6	100:100:4	100:2	6.25

Interpret each profile in STRmix™ and compare the deconvolution against a database containing the profiles of the known donors and non-contributors, assigning an *LR* for each database individual. Profiles of non-contributors can be simulated using allele frequency information for the population of interest. An MS Excel spreadsheet is available from the STRmix™ Technical and Scientific Support platform[7] and can quickly simulate a large

[7] https://support.strmix.com (if necessary please navigate to: Solutions/Implementation and Validation/Implementation assistance/Random profile maker)

number of non-contributor profiles. A database search can be automatically carried out following deconvolution by ticking the 'Perform Database Search' box in the Interpretation screen (refer to the screenshot below).



Note that default search parameters must first be set within the Administration module of STRmix™. These can be found within the Database Search tab of the STRmix™ Defaults window. The *LR* threshold should be set to 0, which will return *LR*s for every individual in the database. Theta (θ) and the option to assign a sub-source *LR* (rather than a sub-sub-source *LR*) can be enabled, if required. Finally, a default priors method can be set for use when assigning a varNOC (variable number of contributors) *LR*. We recommend that the settings be changed to align with those used when interpreting casework profiles (note that if you have elected to model theta using a beta distribution, STRmix™ will use the mean of the distribution when assigning database search *LR*s). Furthermore, we recommend that the allele frequencies used correspond with those used to generate the non-contributor profiles.

When assigning a database search *LR*, STRmix™ considers the following propositions:

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

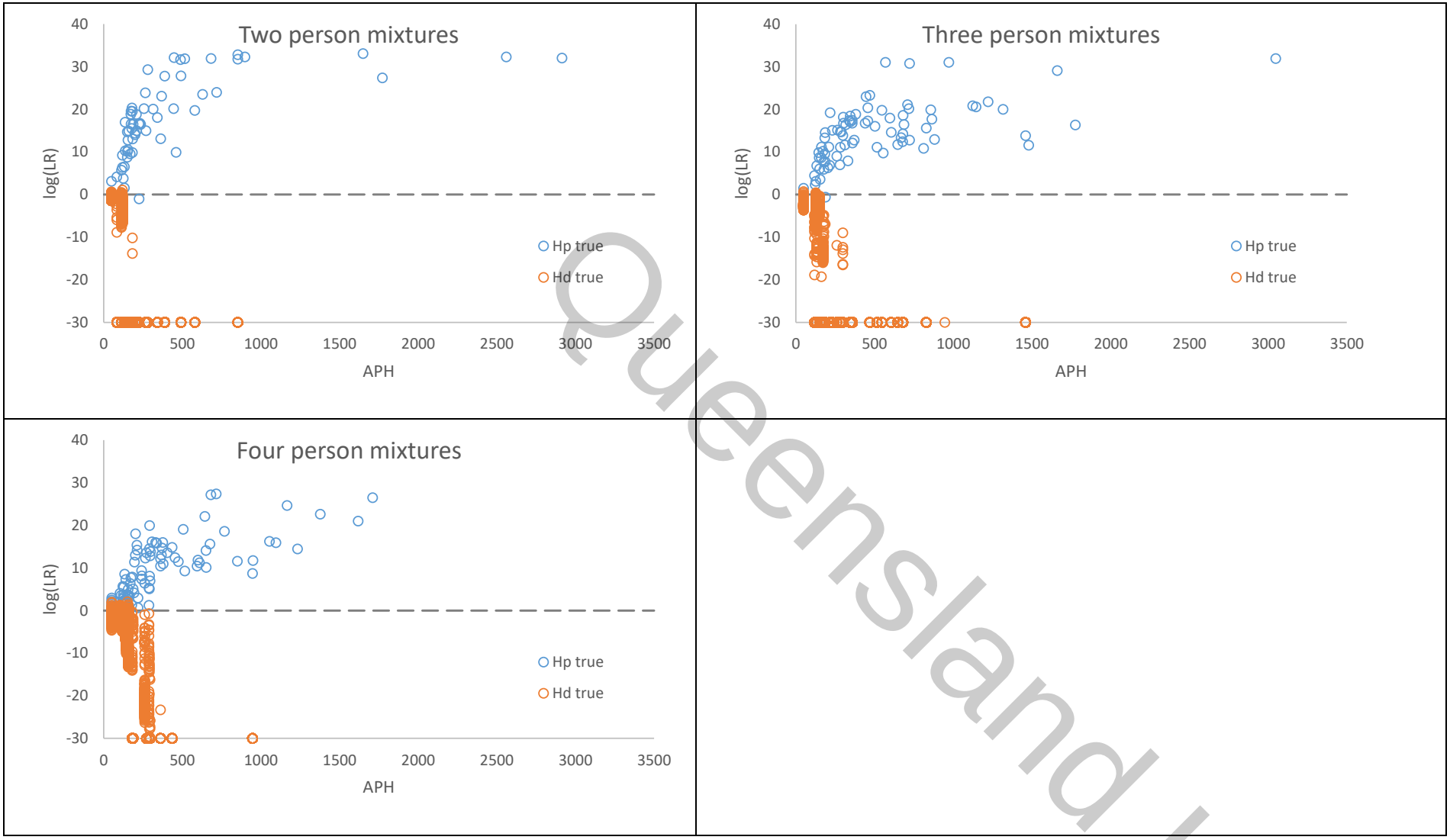
H_d: The DNA originates from N unknown individuals

In the above propositions, N refers to the number of contributors assigned. The true number of contributors to a questioned profile is always unknown. In order to ensure the results are relevant to casework, we recommend having a trained analyst assign N using your laboratory methods and procedures. The assignment of the number of contributors to a profile is complicated by allele sharing, artefacts such as stutter and pull-up, and peaks

below AT. The effect of an incorrect assignment of the number of contributors to an interpretation is explored in 3.7 Section F Effect of number of contributor values on the *LR*.

Plot the log(*LR*) values against the average peak height (APH) per contributor for the two-, three-, and four-person mixtures (plot each number of apparent *N* separately). Exclusions (*LR* = 0) may be plotted using a suitably small number (e.g. log(*LR*) = -30 or -40). APH per contributor may be calculated using unmasked and unshared alleles. Where APH is unable to be calculated for a contributor (e.g. due to dropout or the absence of unmasked/unshared alleles), we suggest instead assigning a value of half AT for APH. APH for non-contributors to a given profile is taken as the minimum APH among the known donors to the profile. Exemplar specificity and sensitivity plots are provided below.

The aim of the test is to explore the limits of STRmix™, particularly the lower limit of DNA where false exclusions (H_p true $LR < 1$) and false inclusions (H_d true $LR > 1$) may arise. The addition of more relevant information (such as increased DNA template, the use of conditioning profiles, and/or the use of PCR replicates) has been shown to improve the performance of STRmix™.



3.6 Section E - Variable propositions

This section covers the following recommendation:

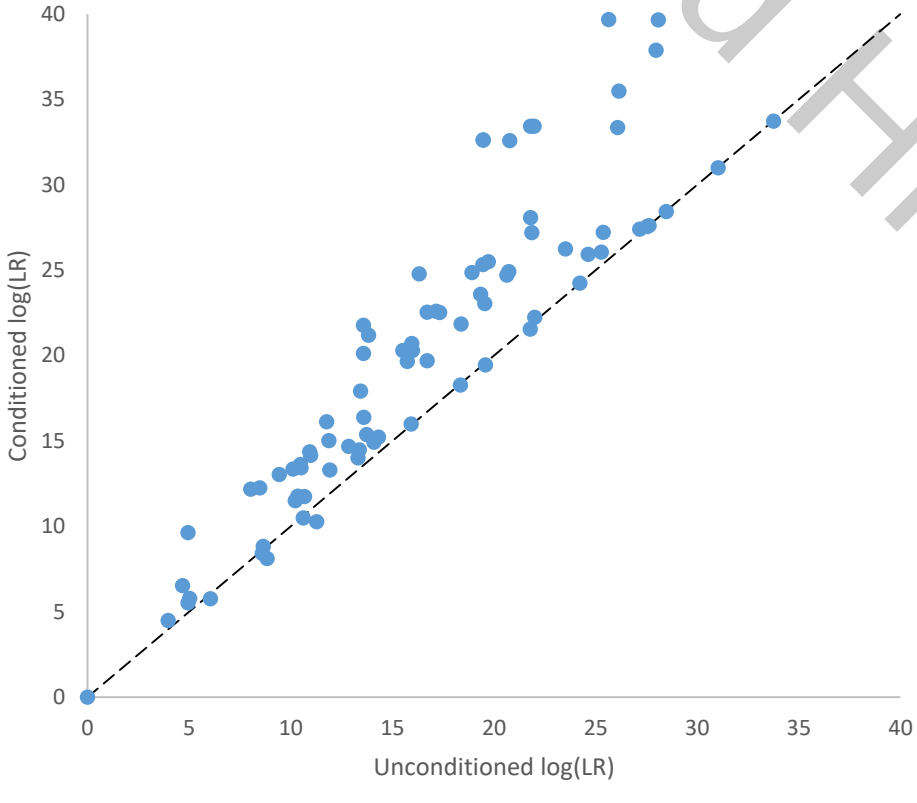
> 4.1.2.1. The laboratory should evaluate more than one set of hypotheses for individual evidentiary profiles to aid in the development of policies regarding the formulation of hypotheses. For example, if there are two persons of interest, they may be evaluated as co-contributors and, alternatively,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

Interpret a selection of profiles from Section D using alternative propositions. For example, repeat the interpretation and condition on one of the known contributors. As before, select 'Perform Database Search' to search the deconvolution against the database used in Section D. The database search settings chosen should correspond with those used previously.

In this example, the different propositions being considered are:

> H_p: The DNA originates from the conditioned individual, the database individual, and N-2 unknown individuals
>
> H_d: The DNA originates from the conditioned individual and N-1 unknown individuals

Compare the log(LR) assigned with the value calculated for the same profile in Section D above. An exemplar plot is provided below.



Values above the dashed line at $x=y$ indicate that the *LR* increased when a conditioning profile was used. The aim of the test is to show that the addition of more relevant information (such as the use of conditioning profiles) improves the performance of STRmix™.

If required, the experiment can be repeated to evaluate the impact of other hypotheses of interest e.g.

> H_p: The DNA originates from individual 1, individual 2, and *N*-2 unknown individuals
>
> H_d: The DNA originates from *N* unknown individuals

Note that compound propositions such as these cannot be evaluated using the database search function. Instead, use the LR from Previous function within STRmix™.

Ultimately, the propositions considered will depend on the case circumstances and on what the prosecution and defence are saying about the DNA results. Refer to the STRmix™ Technical and Scientific Support platform[8] for further discussion on setting appropriate propositions.

3.7 Section F - Effect of number of contributors estimates on the *LR*

This section covers the following recommendation:

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

The effect of uncertainty in the number of contributors has previously been reported for a number of profiles assuming *N*+1, *N*-1, *N*+2, and *N*-2 assigned contributors, where *N* is the true number of contributors [11-15]. The inclusion of an additional contributor beyond that present in the profile has 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* for the major or minor contributor within the profiles.

The effect is tested by both increasing and decreasing the number of contributors relative to the ground truth known value for *N* (i.e. *N*+1 and *N*-1 trials). It should be remembered that the true number of contributors to casework profiles is always unknown.

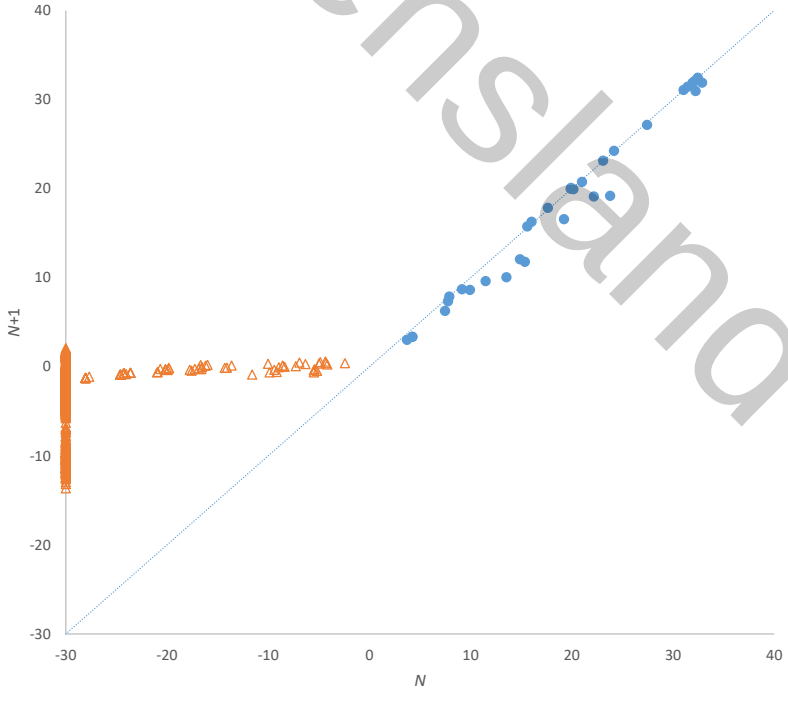
In addition, in an extension of the above method, we recommend testing the *LR* for situations where the POI is not present – that is, when their template is 0. This can be done by using the Mx priors function (refer Section 4.4) and setting the additional contributor's proportion at 0. Re-interpret a subset of the equal mixture proportion samples from Section D (e.g. samples where the DNA amount of the smallest contributor was 200 pg, 100 pg, and 50 pg), increasing the number of contributors by one and using

[8] https://support.strmix.com (if necessary please navigate to: Solutions/Implementation and Validation/Likelihood ratios/The theory of likelihood ratios)

M_x priors with a mean of 0 for the additional contributor position to create the mixture set below. Calculate an *LR* to both known and non-contributors as per Section D using the Database Search function. The expectation would be that the *LR* was inconclusive for non-contributors aligning with the lowest contributor where template ~ 0.

3.7.1 Addition of one contributor

Analysts are likely to add contributors in the presence of ambiguous peaks such as artefact peaks or inflated stutter peaks. Interpret a selection of mixtures from Section D assigning *N*+1 contributors (where the original assigned number of contributors was *N*). As before, select 'Perform Database Search' to search the deconvolution against the database used in Section D. The database search settings chosen should correspond with those used previously. Compare the *LR*s assigned for the known contributors and non-contributors under the assumption of *N* and *N*+1 contributors. An exemplar plot is provided below. In this plot, *LR*s assigned for known donors (H_p true) have been plotted as blue circles whilst *LR*s assigned for non-contributors (H_d true) have been plotted as orange triangles.



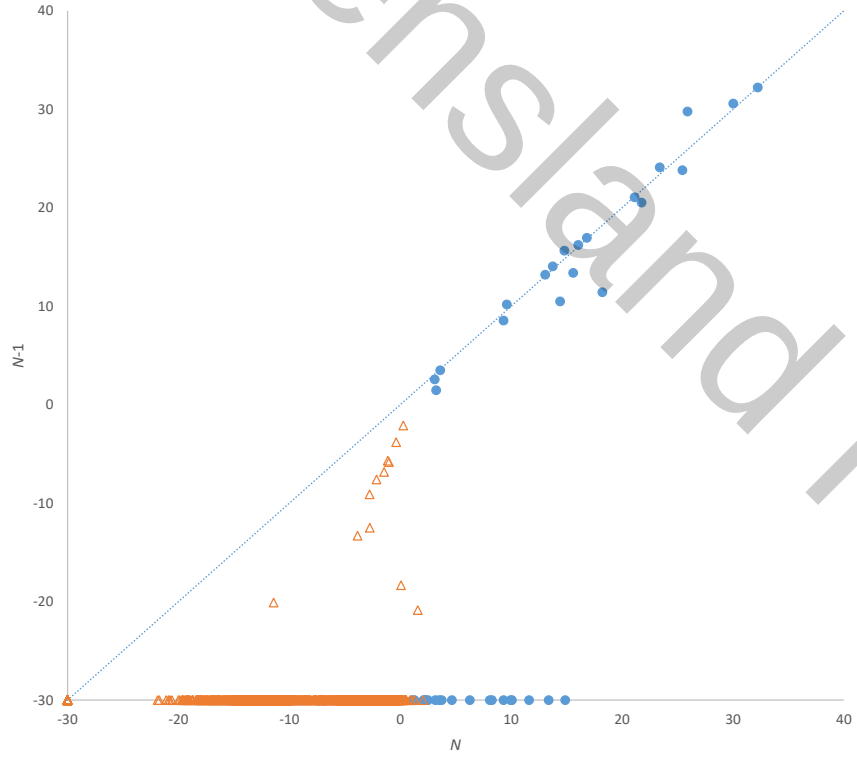
The plot above demonstrates that there was no significant effect on the *LR* for the major or minor contributor when the number of contributors was overestimated. A greater effect would be expected for trace contributors, whose *LR*s may decrease by several orders of magnitude. The inclusion of an additional contributor beyond that present in the profile had the effect of increasing the log(*LR*) for H_d true comparisons. This is the expected result and is due to the addition of a superfluous trace contributor at low template under the assumption of *N*+1 contributors. We expect many genotypes to be accepted for this contributor, including allele and locus dropout. While some low-grade adventitious matches were observed, overestimating the number of contributors did not result in large inclusionary *LR*s for non-contributors.

3.7.2 Subtraction of one contributor

The assumption of one fewer contributor than actually present may be made when contributors are at trace levels and dropping out or where the DNA originates from individuals with similar genotypes (e.g. in family scenarios where mixtures of related individuals might reasonably be expected to be recovered).

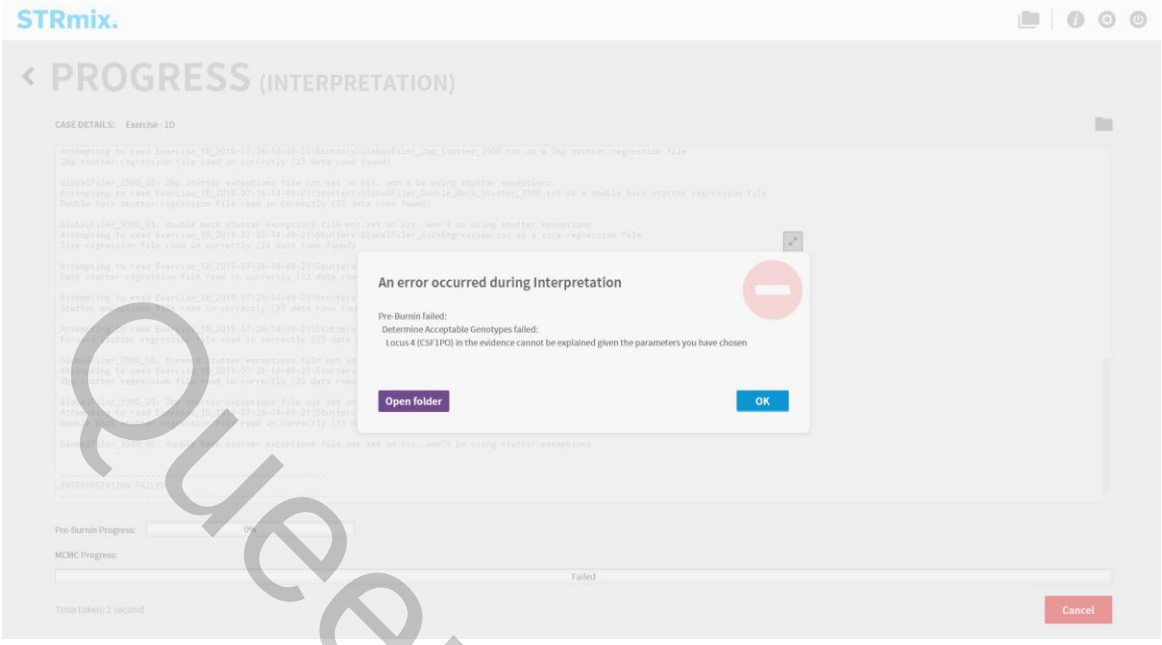
Select a number of mixtures from Section D that, by allele count alone, could be explained as originating from *N*-1 contributors. Interpret the mixtures assuming *N*-1 contributors and search these against the database used in Section D. The database search settings should correspond with those used previously.

Compare the *LR* for the known donors and non-contributors under the assumption of *N* and *N*-1 contributors. An exemplar plot is provided below. In this plot, *LR*s assigned for known donors (H_p true) have been plotted as blue circles whilst *LR*s assigned for non-contributors (H_d true) have been plotted as orange triangles.



As can be seen from the plot above, underestimating the number of contributors to a mixture may lead to false exclusions of true contributors. In general, there is no significant effect on the *LR* of the major or clear minor contributor(s) to the mixture. Rather, the weakest contributor tends to be the one who is falsely excluded.

Underestimating the number of contributors typically results in *LR*s for non-contributors that provide greater support for exclusion. This may be considered to be a conservative outcome. Note that STRmix™ will not progress an interpretation under the assumption of *N*-1 contributors if there are peaks present within the input file that cannot be explained as stutter or drop-in. In this case, STRmix™ will alert the user that the profile cannot be explained by the chosen number of contributors (see screenshot below).



3.8 Section G - Drop-in

This section covers the following recommendation:

4.1.8. Allele drop-in

Assuming your lab has observed drop-in and modelled suitable drop-in parameters, add a realistic drop-in peak to a number of single-source profiles. This can be done *in silico* by modifying the evidence input files used by STRmix™. Interpret as a single-source profile. The *LR* should not change provided that the drop-in peak is within the modelled parameters. Add a peak with height outside of the parameters (i.e. greater than the maximum allowed height for drop-in). The interpretation should not progress as the profile can no longer be explained as originating from one contributor.

3.9 Section H - Stutter

This section covers the following standard:

4.1.9. Forward and reverse stutter

Within STRmix™, stutter is modelled per locus and per allele. From STRmix™ V2.6, any type of stutter product may be modelled. For example, a laboratory may elect to model double back and half back stutter in addition to forward and reverse stutter. Depending on the observations made during the implementation phase, the stutter models developed may be based on:

- Allelic designation,
- LUS designation,
- Per allele average observed *SR*, or
- Per locus average observed *SR*.

When analysing profiles to be used in the validation study, stutter filters should be turned off and labels retained for all stutter variants being modelled. Review select profiles interpreted within Section D and check that stutter modelling is intuitive and in line with expectations. We suggest including profiles where the minor contributor(s) has (have) peak heights at similar levels to stutter peaks from the major contributor(s).

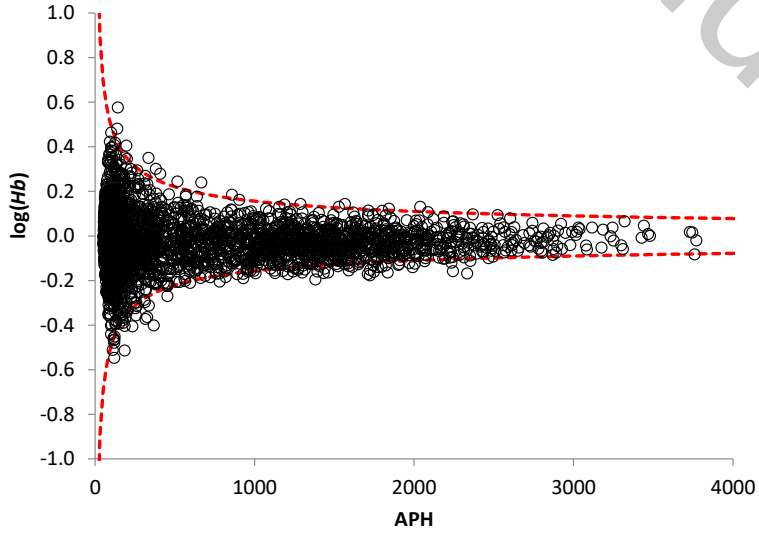
3.10 Section I - Intra-locus peak variance

This section covers the following recommendation:

4.1.10. Intra-locus peak height variance

STRmix™ models peak height variability of single peaks. Laboratory-specific parameters are determined using the Model Maker function of STRmix™ with empirical data.

Traditionally, forensic biologists investigated heterozygote balance (*Hb*), which may be thought of as a measure of peak height variability between the two alleles of a heterozygous locus. A plot of log(*Hb*) versus the average peak height (APH) of a locus demonstrates that variability in *Hb* increases as APH decreases. We can check the performance of the variance models developed by plotting the bounds informed by the Model Maker results (refer to section 2.5 Peak height variance and LSAE variance parameter for further details). A spreadsheet is available from the STRmix™ Technical and Scientific Support platform to assist with this task[9]. An exemplar plot is shown in the following figure.



3.11 Section J - Inter-locus peak variance, DNA degradation and Inhibition

This section covers the following recommendations:

[9] https://support.strmix.com (if necessary please navigate to: Solutions/Implementation and Validation/Implementation assistance/MM data to APH and Hb spreadsheet)

4.1.7. Partial profiles, to include the following:

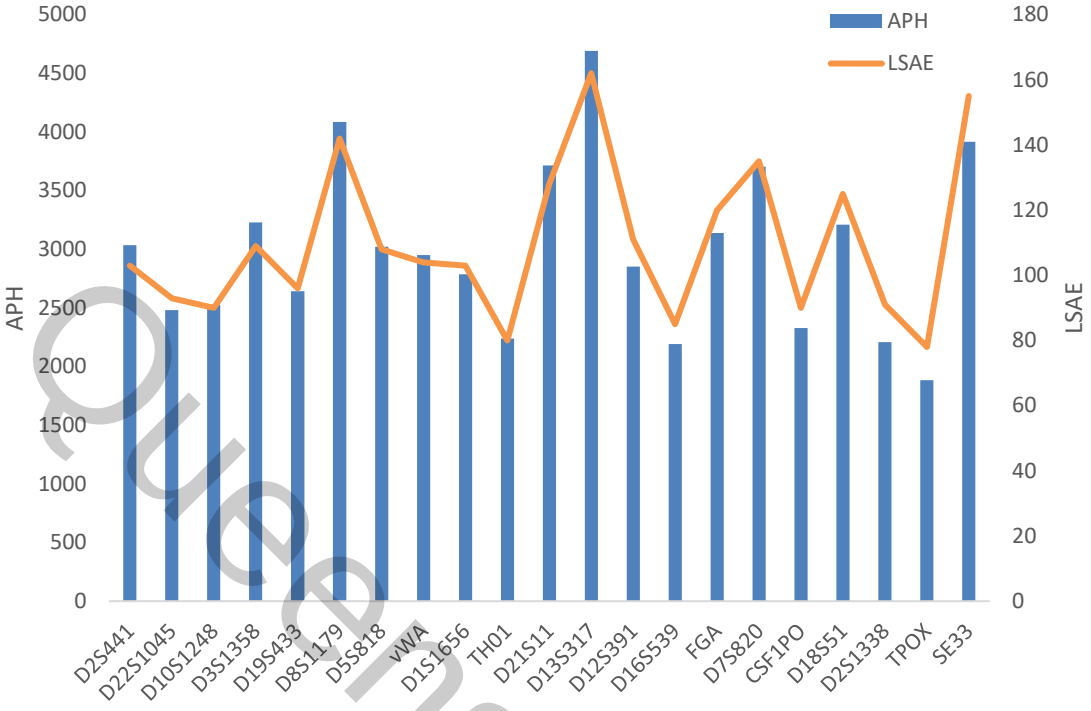
4.1.7.2. DNA degradation

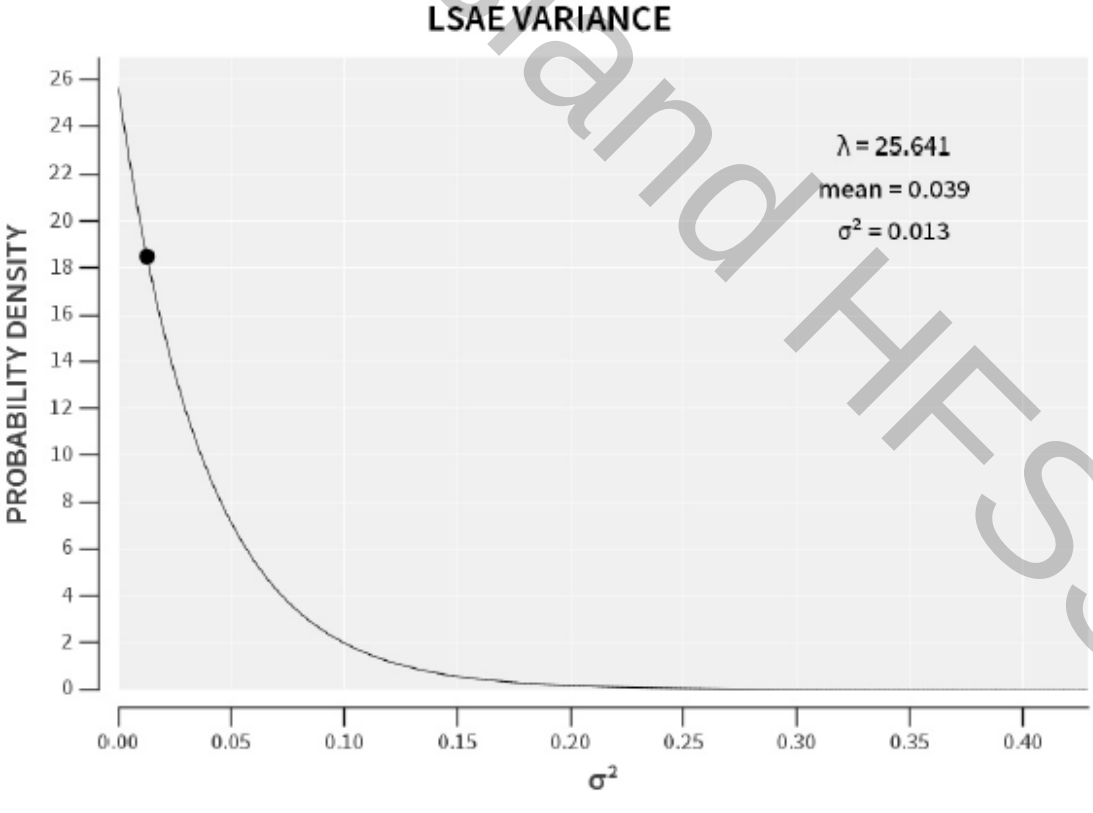
4.1.7.3. Inhibition

4.1.11 Inter-locus peak height variance

Inter-locus peak height variance is modelled within STRmix™ using locus specific amplification efficiencies (LSAE). LSAE is one of the mass parameters considered by STRmix™ during deconvolution and accounts for the observation that even after template and degradation are considered, peak heights at a locus may fall above or below the general trend. As with peak height variance, the prior distribution of LSAE variance is determined using the Model Maker function with empirical data. During interpretation, STRmix™ will optimise LSAE variance and provide the user with the posterior mean LSAE variance parameter which may be used as a diagnostic of the interpretation. Per locus LSAE values are also included within the STRmix™ results following interpretation.

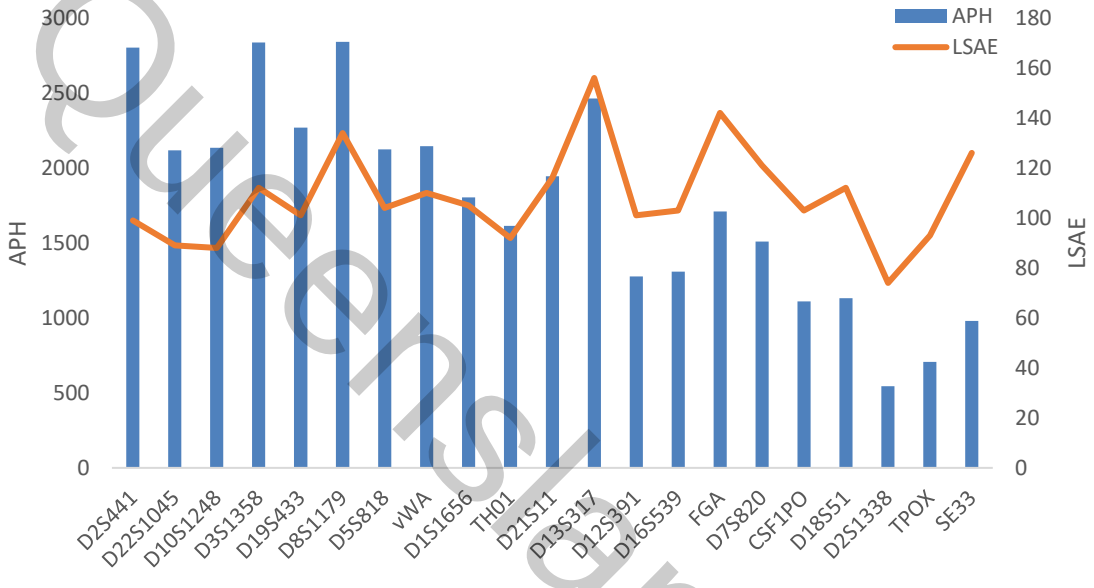
The relationship of LSAE values to average peak height (APH) can be demonstrated via a simple plot. Interpret a single-source profile in STRmix™. Calculate the APH for each locus and collect the LSAE values from the STRmix™ deconvolution. Plot the LSAE and APH per locus as demonstrated in the plot below. This example has been prepared using a single-source GlobalFiler™ profile, with loci arranged in order of increasing molecular weight. The LSAE values observed should roughly correspond with the APH values. As an additional check, the posterior LSAE variance value output by STRmix™ should be typical for profiles generated within your laboratory (i.e. within the body of the LSAE variance prior distribution).

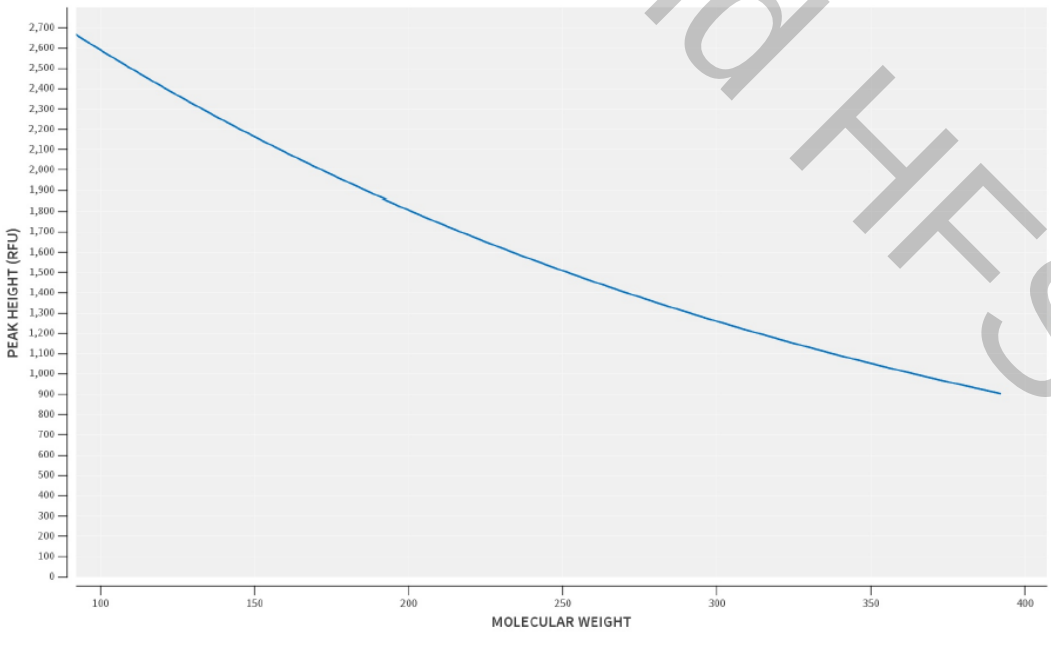


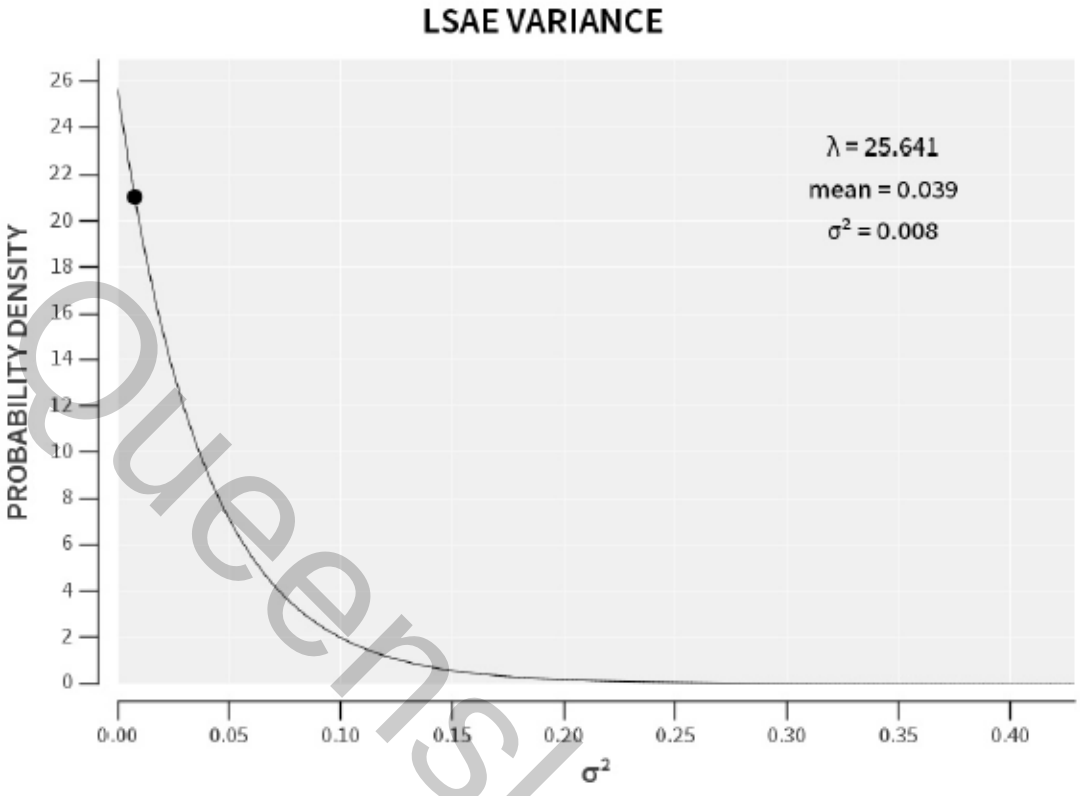


The same single-source input as used in the LSAE experiment above can then be artificially degraded by reducing the peak heights *in silico*. For example, reduce peak heights at the high molecular weight loci by 80% and at the low molecular weight loci by 5%. Interpret the profile within STRmix™, then plot APH and LSAE as before. An exemplar plot is provided below. The effect of degradation can be observed by a

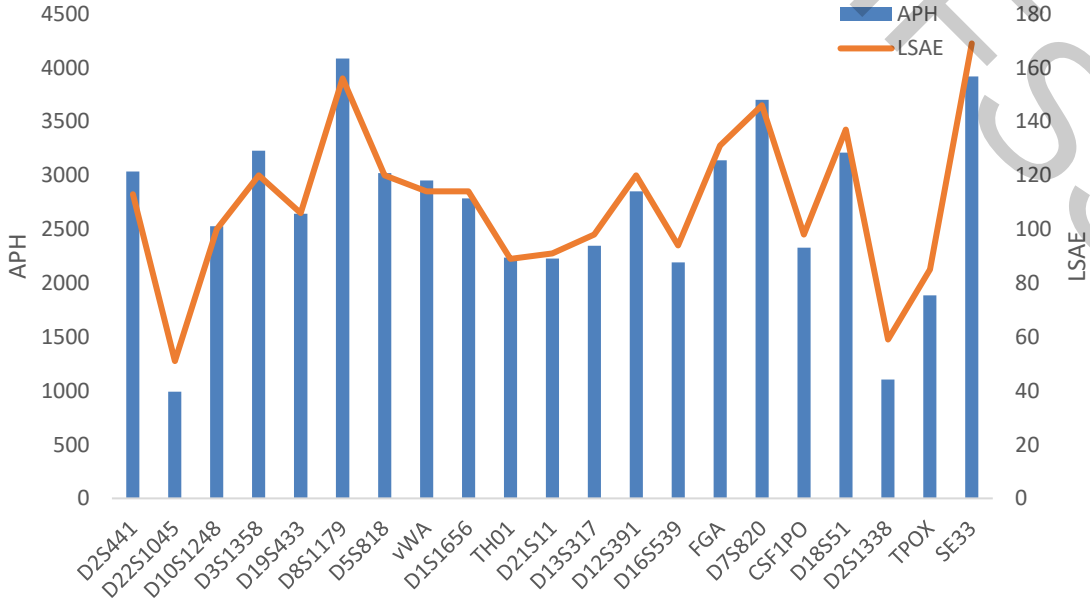
reduction in APH as molecular weight increases. As LSAE is independent of degradation, the LSAE values observed should generally correspond with those seen in the original interpretation using the unedited profile. You should also note an increase in the per contributor degradation value output in the STRmix™ report. This will be evident by a steeper degradation curve for the contributor. As a final check, review the LSAE variance plot. Again, the posterior LSAE variance parameter output by STRmix™ should be typical for profiles generated within your laboratory.

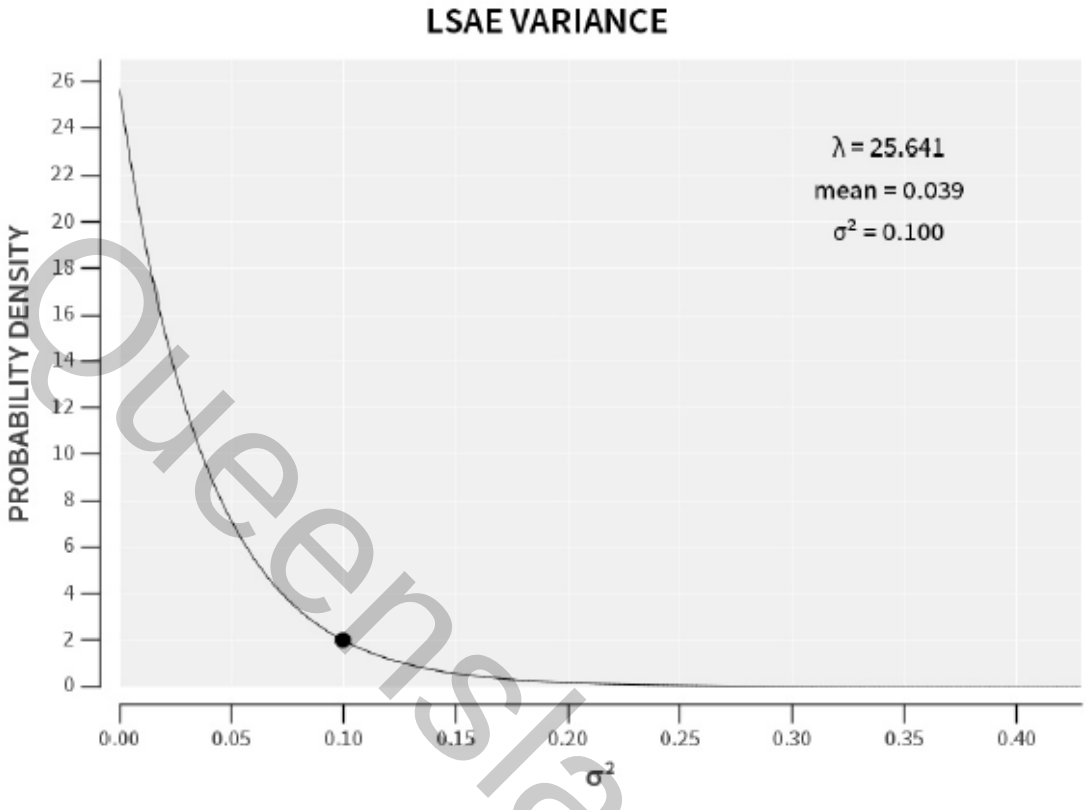






Finally, using the same single-source profile as above, "inhibit" some of the loci by reducing their peak heights *in silico*. Re-interpret using STRmix™, then plot APH and LSAE as before. A decrease in both APH and LSAE should be observed at the inhibited loci. An exemplar plot is provided below; in this example, peak heights at the D22S1045, D21S11, D13S317, and D2S1338 loci were reduced. As a further check, the posterior LSAE variance parameter reported by STRmix™ should be elevated as indicated in the plot below.





3.12 Section K - Additional challenge testing

This section covers the following recommendation:

> 4.1.14. Additional challenge testing (e.g., the inclusion of non-allelic peaks such as bleedthrough and spikes in the typing results)

STRmix™ requires that only numeric values for allele calls are included 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 and an error warning message will be displayed. Note that the use of symbols such as '<' or '>' is also not supported. The presence of a non-allelic peak that has sized within an allelic bin position and is retained within the input file can lead to a number of different outcomes depending on the result observed. These include:

- An exclusion (*LR* = 0). A false exclusion may be observed if the artefact peak is modelled as an allelic peak originating from the contributor of interest (for example if the peak is of a similar height to alleles corresponding to the contributor of interest in a mixed DNA profile).
- No effect. There may be little or no effect on the interpretation if the artefact peak is modelled as drop-in. This requires that drop-in is observed within the laboratory and (i.e. it is being modelled within STRmix™) and that the height of the artefact peak is below the drop-in cap.
- Failure to interpret. STRmix™ may fail to progress an interpretation if the artefact peak increases the minimum number of contributors required to explain the

observed profile. For example, suppose the artefact peak had been recovered at a heterozygous locus in a single-source profile with height exceeding the drop-in cap. STRmix™ will not be able to progress an interpretation under the assumption of a single contributor in this instance, and an error message will be displayed.

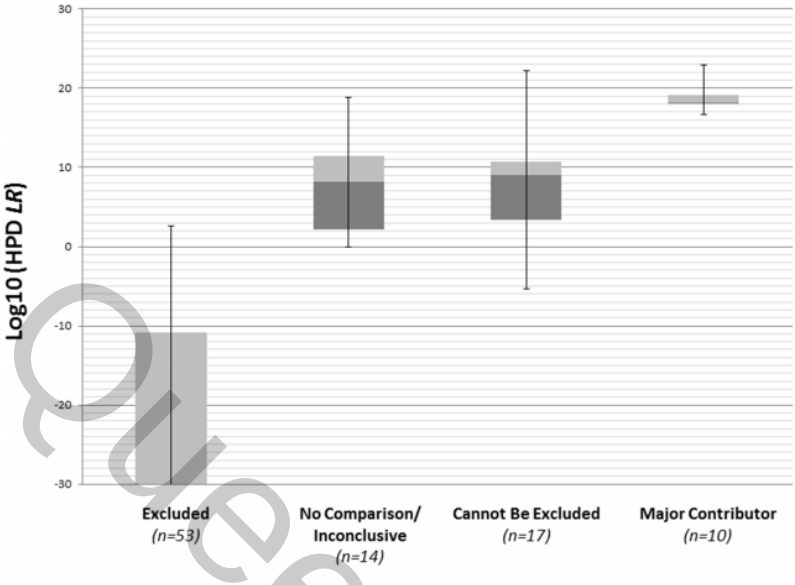
Each of these outcomes can be demonstrated by editing a single-source input file, interpreting within STRmix™, and assigning an *LR* for the known contributor (provided that the interpretation progressed to completion).

3.13 Section L - Comparison to previous interpretation methods

This section covers the following recommendations:

> 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 not directly comparable to binary interpretation. The weights of evidence that are generated by these two approaches are based on different assumptions, thresholds and formulae. However, such a comparison should be conducted and evaluated for general consistency.
>
> > 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.

Review previously interpreted profiles. These could be profiles from proficiency tests, non-probative casework profiles, or profiles used in amplification kit validations. Select a number of profiles covering the range of scenarios where the person of interest (POI) was considered to be excluded, inconclusive, or included. Re-interpret the profiles within STRmix™ and assign an *LR* for the POI. Assess the results to determine if they are intuitive and consistent with the original interpretations. An exemplar plot has been provided below demonstrating one approach to presenting the findings [14].



3.14 Section M - Precision

This section covers the following recommendation:

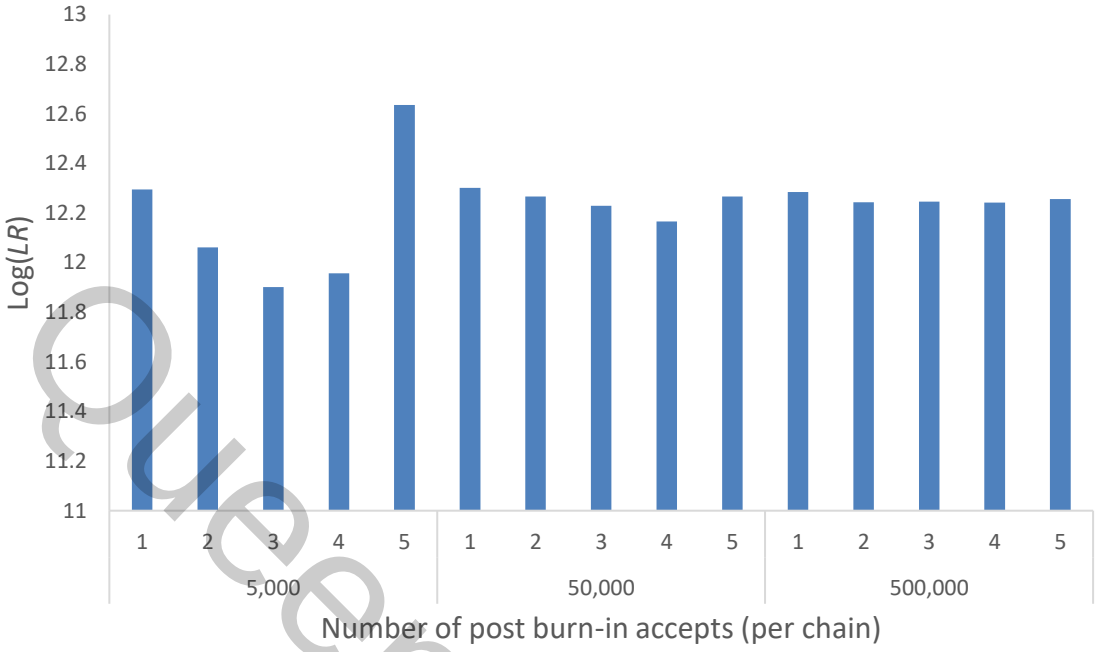
> 4.1.13. Sensitivity, specificity and precision, as described for Developmental Validation

Refer to Section D above for details regarding sensitivity and specificity tests.

STRmix™ uses the Markov chain Monte Carlo (MCMC) process to generate genotype weights during interpretation. This is a random sampling procedure and therefore the weights will vary slightly if the interpretation is repeated. As the weights are used within any subsequent *LR* calculation, the *LR*s assigned from repeat interpretations of the same profile will also vary slightly. The exception to this is for unambiguous single-source profiles or major components where a single genotype has been accepted at each locus with a weight of 1. Typically, the level of variation observed is less than an order of magnitude and is unlikely to be large enough to change the general conclusions drawn from the *LR*.

Select one of the mixed DNA profiles from Section D above where there is considerable genotype uncertainty (for example, a 1:1:1:1 mixture). Interpret the profile multiple times (say, 5 or 10), varying the number of MCMC accepts per chain (for example, 5,000, 50,000, and 500,000 post burn-in accepts per chain). The 'Run Replicates' feature of Batch Mode can be used to quickly set up replicate interpretations of the same profile; refer to the STRmix™ V2.8 Operation Manual for instructions on how to use this feature.

For each deconvolution, assign an *LR* for one of the known contributors and plot the resulting *LR*s. An exemplar plot is provided below. It is expected that as the number of MCMC accepts per chain increases, so too does the precision of STRmix™.



4 INTERNAL VALIDATION OF REPLICATES, MULTI-KITS, VARIABLE NUMBERS OF CONTRIBUTORS, *Mx* PRIORS, MIX TO MIX MATCHING AND THE TOP-DOWN APPROACH

Suggested guidelines for the internal validation of a number of features within STRmix™ are not present in the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems. These features include the use of PCR replicates, multi-kit interpretations, the variable number of contributors (varNOC) function, the mixture proportion (M_x) priors feature, mix to mix matching, and the top-down approach. The use of these features is optional. Suggested approaches for the internal validation of each feature have been provided below.

4.1 Replicates

STRmix™ allows for the interpretation of multiple PCR replicates in the one interpretation, even if different amounts of template DNA have been added to the PCRs. Within STRmix™, the term replicate refers to repeat amplifications of the same DNA extract; multiple injections of the same amplified product are not considered to be replicates and should not be combined within the one interpretation. The use of replicates has been shown to improve the ability of STRmix™ to differentiate true donors from non-contributors, typically increasing *LR*s for true donors and decreasing *LR*s for non-contributors [16].

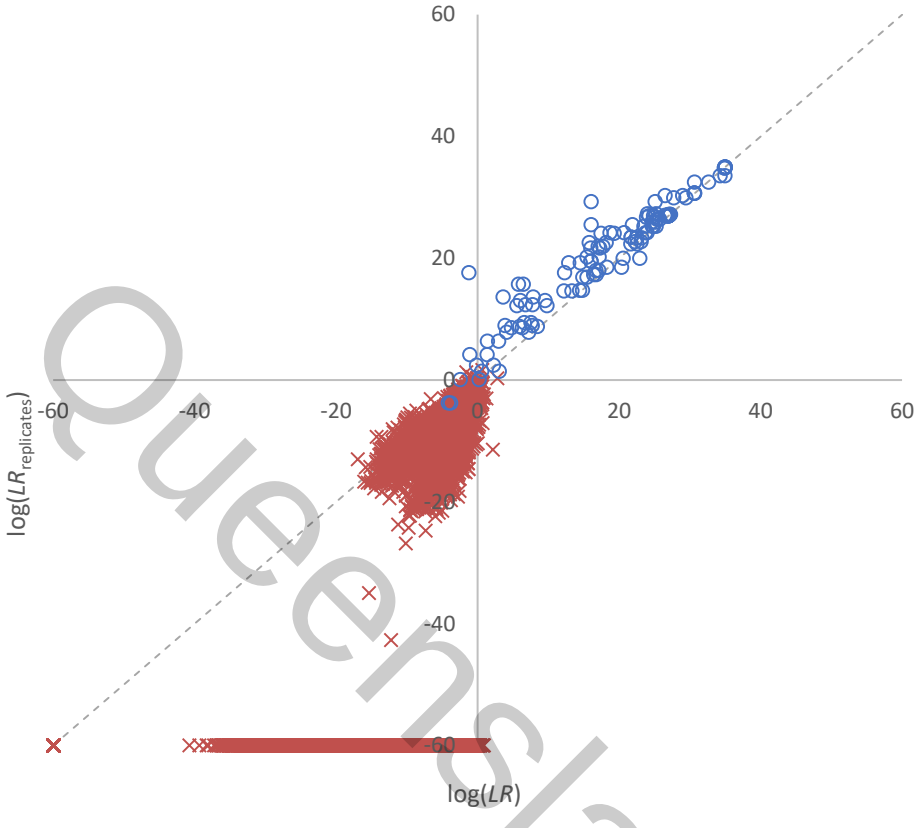
There is a convincing argument against splitting the DNA extract to allow for multiple amplifications [17]. However, assuming sufficient extract remains after the initial amplification, there have been numerous reports of two replicates increasing the information content relative to interpretation using a single amplification, thus providing more data for making inferences about the genotypes of the donor(s) [16, 18-20].

The purpose of this experiment is to determine the effect of the use of PCR replicates on the interpretation. Assuming replicate amplifications were undertaken as part of the validation plan (Section D), re-interpret a selection of profiles using both replicates. As in Section D, select 'Perform Database Search' to search the deconvolution against the database prepared previously. The database search settings chosen should correspond with those used previously. The propositions considered are:

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

H_d: The DNA originates from N unknown individuals

For each mixture, compare the log(*LR*)s assigned when using PCR replicates with those produced from the original interpretation using a single amplification. These may be plotted as shown below. In this example, H_p true *LR*s have been plotted as blue circles whilst H_d true *LR*s have been plotted as red crosses. Data points above the dashed line at $x=y$ indicate an increase in the *LR* when using replicates whilst data points below this line indicate a decrease in the *LR*. In general, we expect to see increased *LR*s for true donors and decreased *LR*s for non-contributors when PCR replicates are used.



4.2 Multi-kits

STRmix™ allows for the use of multiple PCR replicates generated using different profiling kits within the one interpretation. The model assumes that the profiles originate from the same DNA extract. Degradation can either be fixed between kits or allowed to vary per kit. By default, degradation is not fixed. We suggest fixing degradation if the amplifications are carried out concurrently.

The purpose of this experiment is to determine the effect of the use of PCR replicates generated using different profiling kits on the interpretation. Interpret a selection of DNA samples that have been amplified using different kits. As in Section D, select 'Perform Database Search' to search the deconvolution against the database prepared previously. The database search settings chosen should correspond with those used previously. The propositions considered are:

H_p: The DNA originates from the known contributor and N-1 unknown individuals

H_d: The DNA originates from N unknown individuals

As detailed in Section 4.1 above, plot the log(LR)s assigned using multi-kit amplifications against those produced from the original interpretations. Similar to the use of PCR replicates from a single kit, we generally expect LRs for true donors to increase when using multi-kit replicates whilst LRs for non-contributors should decrease.

As an extension of this work, repeat a selection of the interpretations above and fix degradation. Compare the interpretation results and corresponding LRs with the multi-kit interpretation where degradation was not fixed. Provided that there is no substantial difference in degradation between the replicate profiles, it is expected that STRmix™ will give similar results.

4.3 Variable number of contributors (varNOC) function

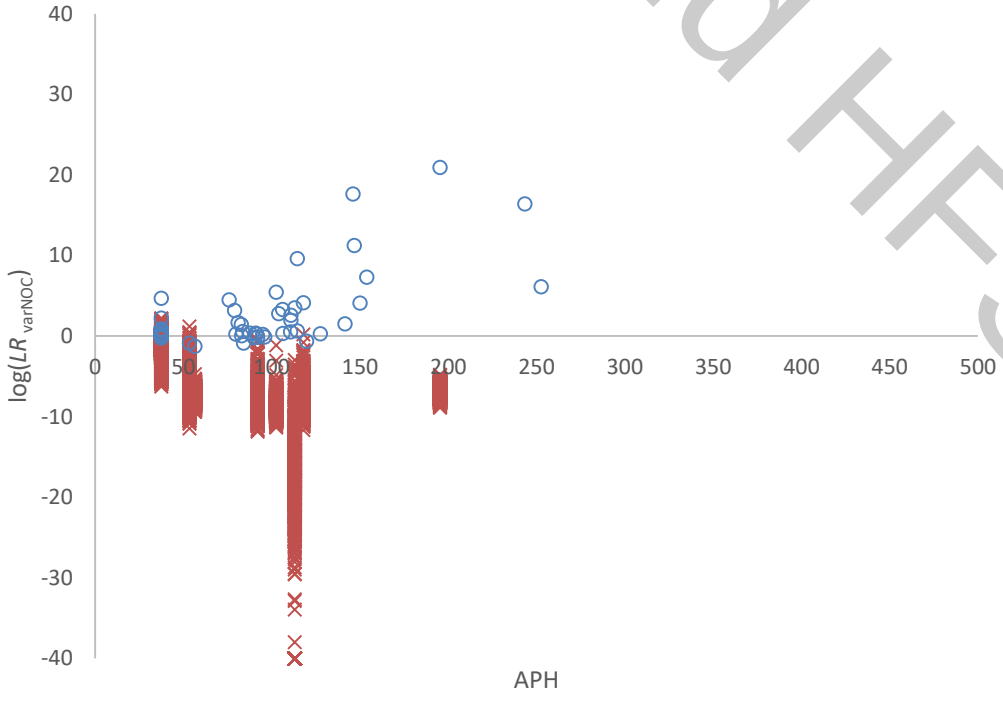
When setting up an analysis in STRmix™, the user is required to input the apparent number of contributors (*N*) to the profile being interpreted. There may be occasions where *N* cannot be assigned with confidence; in these situations, STRmix™ allows for a profile to be interpreted using a range of values for *N*. This feature is referred to as the variable number of contributors (varNOC) function. 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. Whilst any range can be entered, developmental validation of the varNOC function has only been carried out for a contributor range of one [21]. If an increased range is required, it could be argued that too much uncertainty exists in the profile to progress a meaningful interpretation.

The purpose of this experiment is to examine the effect of a varNOC interpretation on subsequent *LR*s. Re-interpret selected profiles from Section D using a contributor range *N*→*N*+1 (where the original assigned number of contributors was *N*). We suggest re-using the profiles interpreted in Section F. When setting up the varNOC interpretations, the 'Population for Range' setting should correpond with the population used to assign the *LR*s in Sections D and F. As in Section D, select 'Perform Database Search' to search the deconvolution against the database prepared previously. The database search settings chosen should correspond with those used previously. We suggest choosing to assign a stratified varNOC *LR* (the default setting). The propositions considered are:

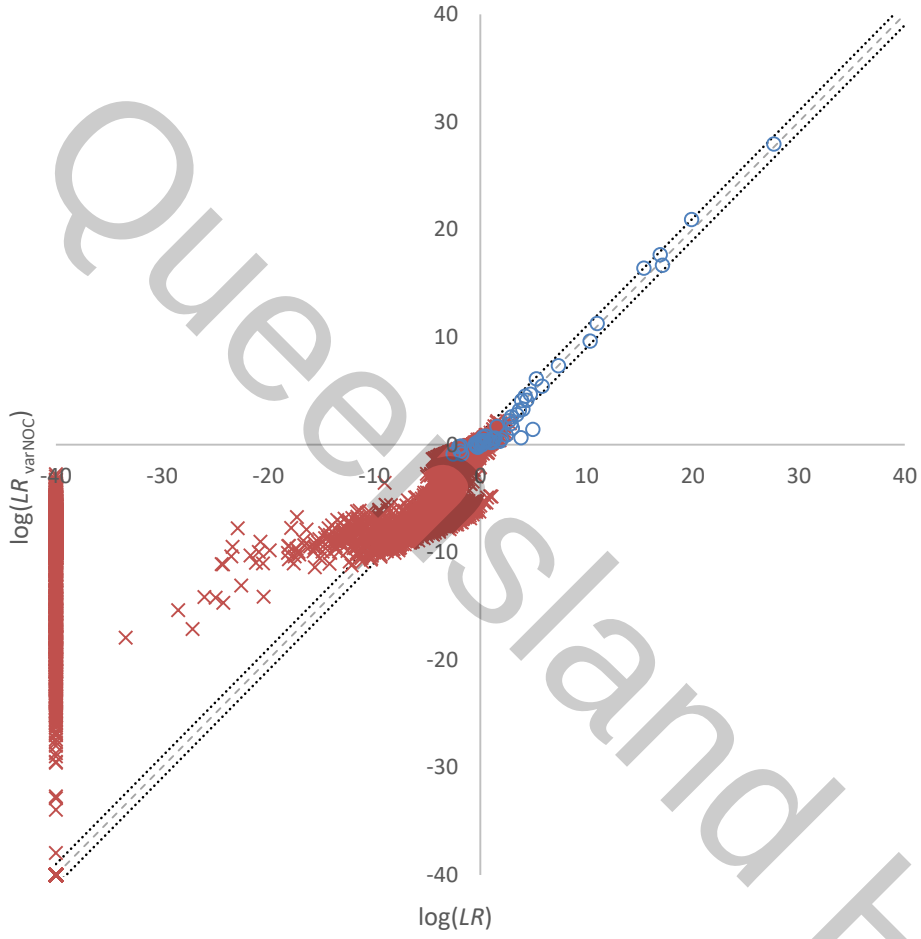
H_p: The database individual is a contributor of DNA to the sample

H_d: The DNA originates from persons who are unrelated to the database individual

As done in Section D, plot the log(*LR*) values against the average peak height (APH) per contributor. Compare the trends in this plot against the corresponding plots from Section D. An exemplar plot is provided below.



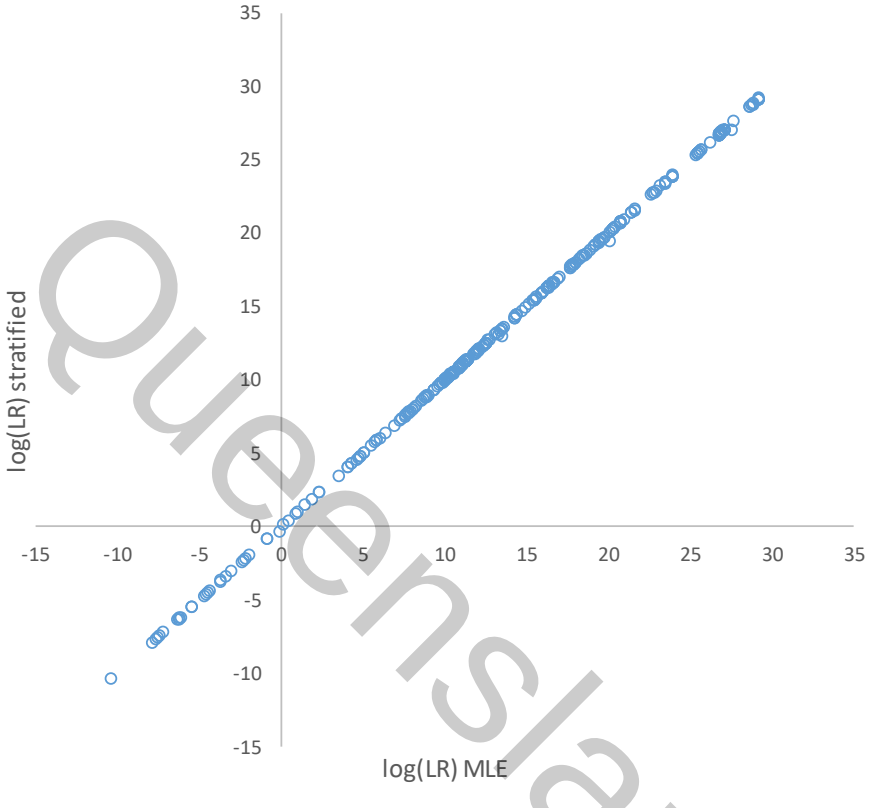
As an additional check, compare the log(*LR*)s for the known donors and non-contributors assigned using the varNOC interpretation (with contributor range *N*→*N*+1) with those generated using the original interpretation (assuming *N* contributors). An exemplar plot is provided below. Within this plot, H_p true *LR*s have been plotted using blue circles whilst H_d true *LR*s have been plotted using red crosses.



This testing may be extended to compare the varNOC *LR* with the results produced under Section F, i.e.:

- Interpret a profile using a range of *N*→*N*+1 and compare with the interpretation assuming *N*+1 contributors from Section F
- Interpret a profile using a range of *N-1*→*N* and compare with the interpretation assuming *N*-1 contributors from Section F

If you are also reporting varNOC *LR*s assigned using the MLE and/or user selected priors methods, repeat this section assigning a varNOC MLE/user selected priors *LR* and then compare the log(*LR*)s produced using the different priors method options (stratify, MLE, and user selected). Provided below is an exemplar plot where varNOC *LR*s assigned using the stratify and MLE priors methods have been compared. In this example, very little difference was observed between the two methods. Note that misuse of the User Selected priors method may result in substantial differences to the *LR* and should be investigated by the laboratory prior to use.



4.4 Mixture proportion (M_x) priors

Occasionally, the mixture proportions reported by STRmix™ may not be intuitive given the observed profile. This may be encountered when interpreting mixed profiles with one or more 'trace' contributors whose peaks are primarily below AT or when interpreting mixed profiles originating from closely-related individuals. Where non-intuitive mixture proportions are reported by STRmix™, re-interpretation using the M_x Priors feature may improve the result.

To use this feature, the analyst must first manually estimate the mixture proportions (M_x). This information is provided to STRmix™ when setting up an interpretation along with an estimate of M_x variance (i.e. how much freedom the user is prepared to give STRmix™ to stray from the chosen proportions). Further details on how to use the M_x Priors feature are provided in the STRmix™ V2.8 Operation manual. Below we provide a suggested approach to validate this feature.

Select a two-person mixture with distinct major and minor contributors from the set used within Section D. Interpret this profile with no M_x priors selected (uninformed). Repeat the interpretation using the M_x Priors feature. A range of values for both mixture proportion and variance should be trialled to investigate the effect on the interpretation. Compare the mixture proportions and log(likelihood) diagnostic reported by STRmix™ across the interpretations carried out. Below we provide an example where a two-person mixture with proportions of approximately 0.80 and 0.20 was interpreted using M_x priors of 0.50 and 0.50, and 0.95 and 0.05. A range of M_x variance values were also explored.

M_x C1 prior	M_x C2 prior	Var C1 prior	Var C2 prior	M_x C1	M_x C2	Log(likelihood)
Uninformed M_x priors				0.80	0.20	24.97
0.50	0.50	0.125	0.125	0.80	0.20	21.17
0.50	0.50	9.766E-04	9.766E-04	0.53	0.47	16.53
0.50	0.50	1.221E-04	1.221E-04	0.50	0.50	16.19
0.95	0.05	1.953E-03	1.953E-03	0.88	0.12	21.58
0.95	0.05	3.052E-05	3.052E-05	0.96	0.04	15.1

It can be seen from the above table that as the variances selected by the user decrease (columns 3 and 4), the mixture proportions reported by STRmix™ (columns 5 and 6) better align with the prior proportions selected by the user (columns 1 and 2). In addition, the log(likelihood) – a measure of how well the expected profile explains the observed profile – decreases. When broad (large) variances are selected (row 2), STRmix™ settles on values similar to those produced when the M_x priors feature is not used (uninformed M_x priors, row 1)

As a further check, examine the interpretations carried out in Section D and identify any profiles where the mixture proportions reported by STRmix™ deviate from the experimental design. Re-interpret these profiles using the M_x Priors feature to investigate whether more intuitive results are able to be obtained.

4.5 Mix to Mix matching

Introduced in STRmix™ V2.7, the Mix to Mix feature allows two or more questioned profiles to be compared in the absence of reference profiles to investigate whether there may be a common contributor between the profiles [22]. There is no restriction on the number of profiles that can be compared or on the number of contributors that each profile can contain (provided that an interpretation within STRmix™ was able to progress to completion). The Mix to Mix feature may be used to provide valuable intelligence information to investigators, potentially linking unsolved cases. Another possible application is to identify sample to sample contamination events during testing [23]. In this section we propose a number of tests that may be carried out to investigate the performance of the Mix to Mix function.

Select a sub-set of mixtures interpreted within Section D. The mixtures chosen should cover a range of contributor numbers, mixture proportions, and template amounts. Ensure that some of the mixtures selected share one or more donors in common with other mixtures in the set. Use the Mix to Mix function within STRmix™ to compare the chosen mixtures. A systematic comparison of each mixture to every other can be achieved by dragging and dropping all of the STRmix™ interpretation run folders to both the left and right hand panes of the Mix to Mix window. Record the average *LR* for each comparison. As the allele frequencies used when assigning a mix to mix *LR* may have a significant effect on the match statistic produced, we recommend that a mix to mix *LR* is assigned using allele frequency data for each of the major sub-populations within your general population and then reporting the lowest average mix to mix *LR* assigned. We

suggest that this approach is also used during validation to ensure that the results obtained are applicable to casework.

Below we provide one approach to presenting the results. Here we have taken six mixtures originating from various donors. Several of these mixtures share one or more donors with one or more of the other mixtures in the set. The mixtures are of varying complexity and total template amount. The Mix to Mix feature was used to compare each mixture with every other, and *LR*s were assigned using four of the FBI extended populations (Caucasian, African American, Southeast Hispanic, and Southwest Hispanic). The minimum average *LR* assigned for each comparison across the four populations is presented in the table below. The number of common donors between the mixtures being compared is indicated in parentheses after the *LR*. Comparisons giving inclusionary *LR*s have been coloured green whilst those producing exclusions (*LR* = 0) or exclusionary LRs have been coloured red.

	9:1 0.31 ng	1:1 0.126 ng	4:4:1 0.567 ng	2:1:1:1 0.625 ng	4:2:1:1:1 0.567 ng	1:1:1:1:1 0.315 ng
9:1 0.31 ng						
1:1 0.126 ng	0 (0)					
4:4:1 0.567 ng	3.1581E-20 (0)	6.0066E17 (2)				
2:1:1:1 0.625 ng	0 (0)	2.8262E3 (1)	2.0997E8 (2)			
4:2:1:1:1 0.567 ng	5.4724E-02 (0)	1.4175E-1 (1)	6.4132E-1 (1)	1.0057E10 (3)		
1:1:1:1:1 0.315 ng	9.9462E-4 (0)	2.6224E10 (2)	2.1808E10 (3)	1.5055E4 (2)	5.1075E0 (1)	

4.6 Top-Down Approach

STRmix™ V2.8 introduces a continuous model variant of the method described by Slooten [24]. This method, called *a Top-Down Approach to DNA Mixtures*, assumes that contributors with a larger relative DNA template contribution are responsible for larger peaks within the profile. Before probabilistic genotyping, analysts would sometimes deduce a major component from a mixture. The Top-Down approach (TDA) can be thought of as a formal method that implements the same ideas. It can be a useful approach to progress the interpretation of high order mixtures, where there is a marked difference in mixture proportion across contributors and only the major components of interest.

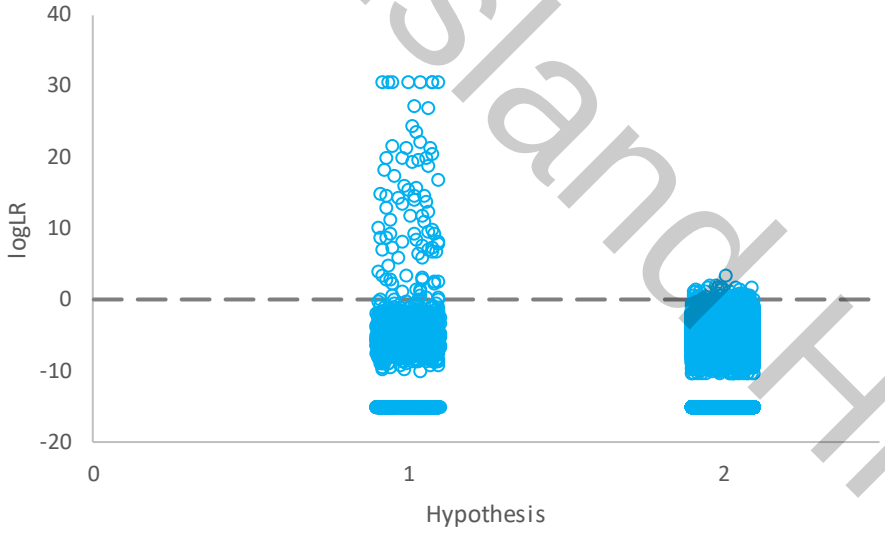
Within STRmix™ V2.8, TDA is implemented by decreasing the per locus analytical threshold (AT) from the maximum observed peak height at the locus to the locus AT described within the STRmix™ kit file, in a defined number of steps (default=20). At each step a sub-profile is generated containing only the peaks at or above the current

per locus AT. Each sub-profile is interpreted and an *LR* assigned for each entry within a database of reference profiles, each taking the role of a POI. The maximum *LR* for each database profile across all the searches is returned as the result of the analysis.

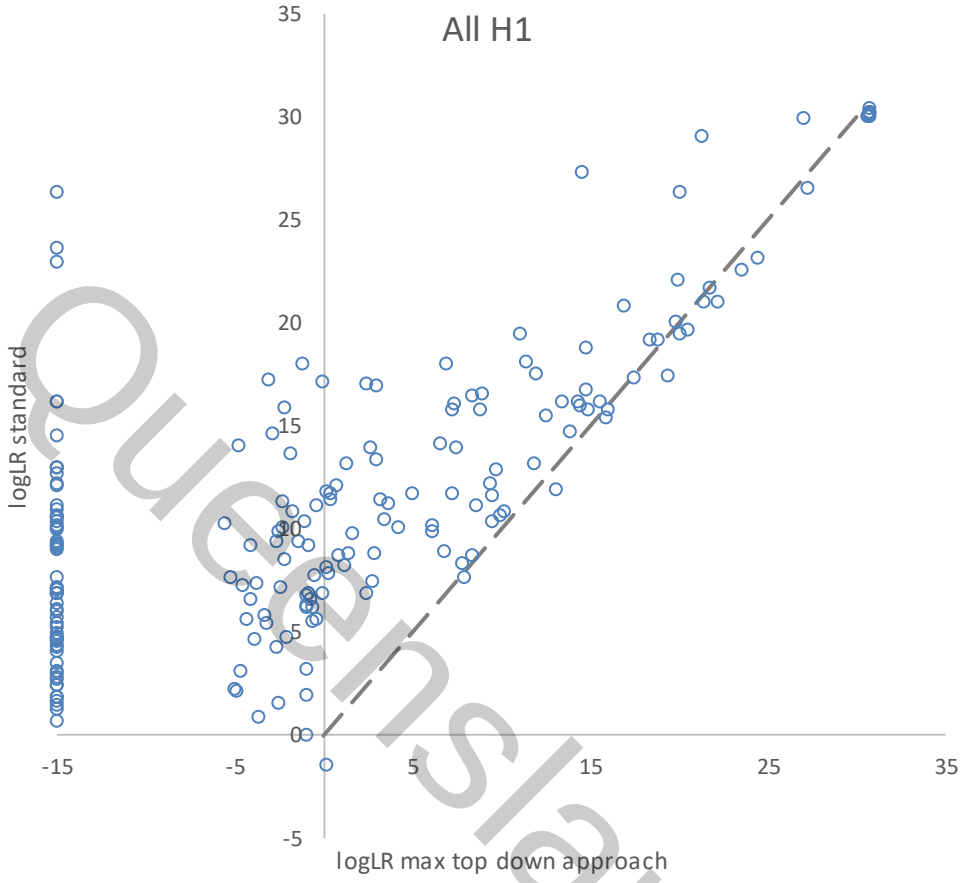
As part of validating this approach we suggest interpreting a set of higher order mixtures, where there is a clear difference between major and minor components. Suitable profiles may be selected from those analysed in Section D. Assign *N* by visual inspection of the profile, where *N* is a subjective assessment of the number of clear major contributors you wish the TDA to consider.

Interpret the selected mixtures in STRmix™ V2.8 using TDA, newly assigned *N,* and use the Database Search function to calculate an *LR* to both known and non-contributors as per Section D.

Assess the range of *LR*s obtained for both known (H1) and non-contributors (H2). In the example given below many of the known contributors become non-contributors under the TDA assigned NOC (i.e. an under-assigned NOC), as expected. In addition, the H1 results can be examined in terms of the posterior mean template amounts per contributor in Section D. The rate of false inclusion of non-contributors (H2) is low, as expected.



For all known contributors plot max TDA log(*LR*) vs log(*LR*) obtained from Section D (an example is given below). Investigate the exclusionary TDA interpretations and false exclusions, including reviewing assigned *N* , the posterior mean template amounts per contributor obtained in Section D, and allele stacking of known contributors.



Examining the TDA step leading to max log(LR) for each interpretation may also be informative.

5 FREQUENTLY ASKED QUESTIONS

1. We have previously validated an earlier version of STRmix™ and wish to move to STRmix™ V2.8. What do we need to do?

The requirements will vary depending on which version you are moving from and what features you wish to use. Please review the relevant Release and Testing reports on the STRmix™ Technical and Scientific support platform[10] and contact support@strmix.com for further information and assistance.

2. We use multiple CE instruments within our lab. Do we need to determine parameters for each instrument?

Provided that the instruments are from the same series (e.g. multiple 3500 instruments), it is expected that the parameters will not differ substantially between individual machines. Significant differences are expected between instruments from different series (e.g. 3130 versus 3500) and individual parameters will need to be determined.

3. We use multiple PCR and injection protocols. Do we need to determine parameters for each protocol?

Yes. Variations to the amplification and/or injection conditions used may affect some of the parameters. Parameters will need to be determined for each kit, PCR protocol, CE instrument, and injection protocol combination. Where a parameter is noted to differ considerably between protocols, separate kit files should be set up within STRmix™. Where no substantial difference is observed, a single set of parameters determined using a combined dataset may be adopted. If in doubt, please contact support@strmix.com for further information and assistance.

4. Our laboratory is part of a wider laboratory system across multiple sites. Do we need to determine parameters for each site?

Parameters may or may not differ between laboratory sites. We recommend that parameters are first determined separately for each site. Where no substantial differences are observed between sites, a combined dataset may be prepared and used to determine parameters that are suitable for use by all sites. If in doubt, please contact support@strmix.com for further information and assistance.

5. What types of stutter should I model?

In STRmix™ V2.6 and later, any type of stutter product may be modelled. We recommend modelling those variants that are routinely observed above your casework AT. While you may elect to model additional variants that are observed above AT on occasion, the collection of sufficient data to inform the models may prove to be challenging (and the modelling of additional stutter variants will also increase interpretation runtimes within STRmix™). One approach to handle such variants is to remove labels from these putative stutter peaks observed above AT during profile analysis.

6. Allele regression, LUS regression, or per allele average: how do I determine which stutter model to select?

[10] https://support.strmix.com (if necessary please navigate to: Solutions/Implementation and Validation/Release and Testing Reports – STRmix™)

Begin by plotting observed stutter ratio versus allelic designation. Add a trend line to the plot, then assess how well the trend line describes the observed stutter ratios. Is there a nice, linear relationship between allelic designation and stutter ratio? Are there any microvariant alleles that fall off the general trend? A classic example of this is the TH01 9.3 allele. For those loci where allelic designation is observed to be a poor predictor of stutter ratio, plot observed stutter ratio versus LUS designation. Add a trend line as before and assess. If neither allele nor LUS designation prove to be reasonable explanatory variables for stutter ratio, we recommend that the average observed stutter ratio is calculated for each allele at the locus. These values can then be input into the stutter exceptions file and used to model expected stutter ratios within STRmix™.

7. Drop-in? Isn't that just contamination?

Drop-in is generally defined as the observation of one or a few low-level peaks within a DNA profile. Drop-in has previously been described as alleles "snowing from the ceiling" and falling into the PCR tube during amplification set-up. Importantly, drop-in is not reproducible if the DNA extract is subsequently re-amplified. In contrast, contamination is reproducible. Furthermore, it may be possible to identify the source of the contaminant DNA. When determining drop-in parameters for use in STRmix™, any negative controls that appear to have been contaminated should be excluded from the dataset.

8. I've used the drop-in calculator available from the STRmix™ support site. How do I determine whether the gamma model developed is suitable?

We recommend examining the fit of the model to the observed data. Where limited observations of drop-in have been made, use of a uniform model is instead recommended. If in doubt, please contact support@strmix.com for further information and assistance.

9. Can STRmix™ implement recommendation 4.1 of NRCII?

No. STRmix™ uses the Balding and Nichols sub-population model (recommendation 4.2 in NRCII). This is the most appropriate population genetic model to use when assigning a match probability and provides a highly conservative estimate (on the order of 99 to 1). All probabilistic genotyping software implement the Balding and Nichols model.

10. STRmix™ falsely excluded the known donor to my profile in Section D. Why?

Possible causes may be due to the under-assignment of the number of contributors, an input file error, or the presence of an unresolved allele [25]. Note that unresolved alleles seperated by 1bp are occasionally observed at some loci (e.g. D1S1656 and TH01) and are a limitation of CE technology. Rework options such as re-injection of the sample may resolve the issue. Otherwise, we recommend that the affected locus is ignored during interpretation. For further troubleshooting advice, please contact support@strmix.com for further information and assistance.

6 REFERENCES

[1] J.-A. Bright, J.M. Curran, Investigation into stutter ratio variability between different laboratories, Forensic Sci. Int. Genet. 13(0) (2014) 79-81.

[2] H. Kelly, J.-A. Bright, M. Kruijver, S. Cooper, D. Taylor, K. Duke, M. Strong, V. Beamer, C. Buettner, J. Buckleton, A sensitivity analysis to determine the robustness of STRmix™ with respect to laboratory calibration, Forensic Sci. Int. Genet. 35 (2018) 113-122.

[3] D. Taylor, J.-A. Bright, C. McGovern, C. Hefford, T. Kalafut, J. Buckleton, Validating multiplexes for use in conjunction with modern interpretation strategies, Forensic Sci. Int. Genet. 20 (2016) 6-19.

[4] J.R. Gilder, T.E. Doom, K. Inman, M. Crim, D.E. Krane, Run-Specific Limits of Detection and Quantitation for STR-Based DNA Typing, Journal of the Forensic Sciences 52(1) (2006) 97-101.

[5] J. Bregu, D. Conklin, E. Coronado, M. Terrill, R.W. Cotton, C.M. Grgicak, Analytical Thresholds and Sensitivity: Establishing RFU Thresholds for Forensic DNA Analysis, Journal of Forensic Sciences 58(1) (2013) 120-129.

[6] U.J. Mönich, K. Duffy, M. Médard, V. Cadambe, L.E. Alfonse, C. Grgicak, Probabilistic characterisation of baseline noise in STR profiles, Forensic Sci. Int. Genet. 19 (2015) 107-122.

[7] D. Taylor, J.-A. Bright, J.S. Buckleton, The interpretation of single source and mixed DNA profiles Forensic Science International: Genetics 7(5) (2013) 516-528.

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

[9] J.-A. Bright, D. Taylor, C.E. McGovern, S. Cooper, L. Russell, D. Abarno, J.S. Buckleton, Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles, Forensic Sci. Int. Genet. 23 (2016) 226-239.

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

[11] J.-A. Bright, J.M. Curran, J.S. Buckleton, The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation, Forensic Sci. Int. Genet. 12 (2014) 208-214.

[12] 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.

[13] J.-A. Bright, K. Cheng, Z. Kerr, C. McGovern, H. Kelly, T.R. Moretti, M.A. Smith, F.R. Bieber, B. Budowle, M.D. Coble, R. Alghafri, P.S. Allen, A. Barber, V. Beamer, C. Buettner, M. Russell, C. Gehrig, T. Hicks, J. Charak, K. Cheong-Wing, A. Ciecko, C.T. Davis, M. Donley, N. Pedersen, B. Gartside, D. Granger, M. Greer-Ritzheimer, E. Reisinger, J. Kennedy, E. Grammer, M. Kaplan, D. Hansen, H.J. Larsen, A. Laureano, C. Li, E. Lien, E. Lindberg, C. Kelly, B. Mallinder, S. Malsom, A. Yacovone-Margetts, A. McWhorter, S.M. Prajapati, T. Powell, G. Shutler, K. Stevenson, A.R. Stonehouse, L. Smith, J. Murakami, E. Halsing, D. Wright, L. Clark, D.A. Taylor, J. Buckleton, STRmix™

collaborative exercise on DNA mixture interpretation, Forensic Sci. Int. Genet. 40 (2019) 1-8.

[14] T.R. Moretti, R.S. Just, S.C. Kehl, L.E. Willis, J.S. Buckleton, J.-A. Bright, D.A. Taylor, A.J. Onorato, Internal validation of STRmix for the interpretation of single source and mixed DNA profiles, Forensic Sci. Int. Genet. 29 (2017) 126-144.

[15] T. Bille, S. Weitz, J.S. Buckleton, J.-A. Bright, Interpreting a major component from a mixed DNA profile with an unknown number of minor contributors, Forensic Sci. Int. Genet. 40 (2019) 150-159.

[16] D. Taylor, Using continuous DNA interpretation methods to revisit likelihood ratio behaviour, Forensic Sci. Int. Genet. 11 (2014) 144-153.

[17] T.W. Bille, M.D. 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, 2016.

[18] S. Gittelson, C.R. Steffen, M.D. Coble, Low-template DNA: A single DNA analysis or two replicates?, Forensic Sci. Int. 264 (2016) 139-145.

[19] B. Kokshoorn, B.J. Blankers, Response to Grisedale and Van Daal: comparison of STR profiling from low template DNA extracts with and without the consensus profiling method, Investigative Genetics 4 (2013) 1-1.

[20] C.D. Steele, M. Greenhalgh, D.J. Balding, Verifying likelihoods for low template DNA profiles using multiple replicates, Forensic Sci. Int. Genet. 13(0) (2014) 82-89.

[21] C. McGovern, K. Cheng, H. Kelly, A. Ciecko, D. Taylor, J.S. Buckleton, J.-A. Bright, Performance of a method for weighting a range in the number of contributors in probabilistic genotyping, Forensic Sci. Int. Genet. 48 (2020).

[22] J.-A. Bright, D. Taylor, Z. Kerr, J. Buckleton, M. Kruijver, The efficacy of DNA mixture to mixture matching, Forensic Sci. Int. Genet. 41 (2019) 64-71.

[23] D. Taylor, E. Rowe, M. Kruijver, D. Abarno, J.-A. Bright, J. Buckleton, Inter-sample contamination detection using mixture deconvolution comparison, Forensic Sci. Int. Genet. 40 (2019) 160-167.

[24] K. Slooten, A top-down approach to DNA mixtures, Forensic Science International: Genetics 46 (2020) 102250.

[25] J.-A. Bright, R. Richards, M. Kruijver, H. Kelly, C. McGovern, A. Magee, A. McWhorter, A. Ciecko, B. Peck, C. Baumgartner, C. Buettner, S. McWilliams, C. McKenna, C. Gallacher, B. Mallinder, D. Wright, D. Johnson, D. Catella, E. Lien, C. O'Connor, G. Duncan, J. Bundy, J. Echard, J. Lowe, J. Stewart, K. Corrado, S. Gentile, M. Kaplan, M. Hassler, N. McDonald, P. Hulme, R.H. Oefelein, S. Montpetit, M. Strong, S. Noël, S. Malsom, S. Myers, S. Welti, T. Moretti, T. McMahon, T. Grill, T. Kalafut, M. Greer-Ritzheimer, V. Beamer, D.A. Taylor, J.S. Buckleton, Internal validation of STRmix™ – A multi laboratory response to PCAST, Forensic Sci. Int. Genet. 34 (2018) 11-24.

7 APPENDIX 1 KEY STRMIX™ REFERENCES

[1] D.A. Taylor, J.-A. Bright, J. S. Buckleton, The interpretation of single source and mixed DNA profiles, Forensic Science International: Genetics. 7(5) (2013) 516-528. **(Core maths paper)**

[2] J.-A. Bright, D.A. Taylor, J. M. Curran, J. S. Buckleton, Developing allelic and stutter peak height models for a continuous method of DNA interpretation, Forensic Science International: Genetics. 7(2) (2013) 296-304. **(Core models paper)**

[3] J.-A. Bright, D.A. Taylor, J. Curran, J.S. Buckleton, Degradation of forensic DNA profiles, Australian Journal of Forensic Sciences. 45(4) (2013) 445-449.

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

[5] D.A. Taylor. Using continuous DNA interpretation methods to revisit likelihood ratio behaviour. Forensic Science International: Genetics, 2014. 11: 144-153.

[6] 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.

[7] J.-A. Bright, I.W. Evett, D.A. 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.

[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.A. 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] T.W. Bille, S.M. Weitz, M.D. Coble, J.S. Buckleton and J.-A. Bright, Comparison of the performance of different models for the interpretation of low level mixed DNA profiles. Electrophoresis, 2014. 35:3125-33.

[11] D.A. Taylor, J.-A. Bright and J.S. Buckleton, The 'factor of two' issue in mixed DNA profiles. Journal of Theoretical Biology, 2014. 363: p. 300-306.

[12] D.A. Taylor and J.S. Buckleton, Do low template DNA profiles have useful quantitative data? Forensic Science International: Genetics, 2015. 16:13-6.

[13] D.A. Taylor, J.S. Buckleton and I. Evett, Testing likelihood ratios produced from complex DNA profiles. Forensic Science International: Genetics, 2015. 16:165-171.

[14] S.J. Cooper, C.E. McGovern, J.-A. Bright, D.A. Taylor and J.S. Buckleton, Investigating a common approach to DNA profile interpretation using probabilistic software. Forensic Science International: Genetics, 2015. 16:121-131.

[15] J.-A. Bright, D.A. Taylor, C.E. McGovern, S.J. Cooper, L.J. Russell, D.V. Abarno and J.S. Buckleton, Developmental validation of STRmix™, expert software for the

interpretation of forensic DNA profiles. Forensic Science International: Genetics, 2016. 23:226-239.

[16] D.A. Taylor, J.-A. Bright, C.E. McGovern, C. Hefford, T. Kalafut, J.S. Buckleton, Validating multiplexes for use in conjunction with modern interpretation strategies. Forensic Science International: Genetics, 2016. 20:6-19.

[17] D.A. Taylor, J.S. Buckleton, J.-A. Bright, Factors affecting peak height variability for short tandem repeat data. Forensic Science International: Genetics, 2016. 21:126-133.

[18] T.R. Moretti, R.S. Just, S.C. Kehl, L.E. Willis, J.S. Buckleton, J.-A. Bright, D.A. Taylor, Internal validation of STRmix™ for the interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics, 2017. 29:126-144.

[19] D.A. Taylor, J.-A. Bright, H. Kelly, M.-H. Lin, J.S. Buckleton. A fully continuous system of DNA profile evidence evaluation that can utilise STR profile data produced under different conditions within a single analysis. Forensic Science International: Genetics, 2017. 31:149-154.

[20] D.A. Taylor, J.S. Buckleton, J.-A. Bright, Does the use of probabilistic genotyping change the way we should view sub-threshold data? Australian Journal of Forensic Sciences, 2017. 49(1):78-92.

[21] J.-A. Bright, et al., Internal validation of STRmix; A multi laboratory response to PCAST. Forensic Science International: Genetics, 2018. 34:11-24.

[22] J.S. Buckleton, et al., The Probabilistic Genotyping Software STRmix: Utility and Evidence for its Validity. Journal of Forensic Sciences, 2019. 64(2): 393-405.

[23] D.A. Taylor, J.-A. Bright , J.S. Buckleton, Interpreting forensic DNA profiling evidence without specifying the number of contributors. Forensic Science International: Genetics. 2014;13:269-80.

[24] D.A. 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.

[25] T. Bille, S. Weitz, J.S. Buckleton, J.-A. Bright, Interpreting a major component from a mixed DNA profile with an unknown number of minor contributors. Forensic Science International: Genetics. 2019;40:150-9.

8 APPENDIX 2 TIPS FOR PERFORMANCE CHECKS

A performance check is a check of the functionality of the software. It may be undertaken in lieu of a full internal validation where changes to the software do not substantially impact interpretation or analysis of the typing results.

A performance check may involve the re-evaluation of data used during a laboratory's initial internal validation. A number of profiles of varying quality and complexity should be identified. It is recommended that these profiles are interpreted by the laboratory in the STRmix™ version currently in casework use and the results used as a benchmark.

When carrying out a performance check of a new version of STRmix™, re-interpret the above set of profiles in the new version and compare with the results generated using the current version. From STRmix™ V2.5 and later, the use of config files can facilitate this task. For more information on how to use config files within STRmix™, please refer to the STRmix™ Technical and Scientific Support platform[11].

Furthermore, when upgrading to a new version of STRmix™, we recommended reviewing all relevant Release and Testing reports to inform the scope of the testing required. These reports are available on the STRmix™ Technical and Scientific Support platform[12].

[11] https://support.strmix.com (if necessary please navigate to: Solutions/STRmix training videos/STRmix V2.6 videos/Config files in STRmix v2.6)
[12] https://support.strmix.com (if necessary please navigate to: Solutions/Implementation and Validation/Release and Testing Reports – STRmix™)

9 APPENDIX 3 SUMMARY OF SWGDAM GUIDELINES FOR THE INTERNAL VALIDATION OF PROBABILISTIC GENOTYPING SYSTEMS

Recommendation	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 and L
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