

# Use of STRmix™ Software

## 1 Purpose

The purpose of this document is to describe the use of STRmix™ for the analysis of DNA profiles generated using the PowerPlex® 21 profiling kit (PP21) and the creation of STRmix™ variance parameters.

## 2 Scope

This procedure applies to all scientists within Forensic DNA Analysis who perform the interpretation of DNA profiling results. Additional information can be obtained from the most current STRmix™ User's Manual.

## 3 Definitions

H<sub>1</sub> (named H<sub>P</sub> in STRmix™): Hypothesis on the numerator of the LR (includes POI)

H<sub>2</sub> (named H<sub>D</sub> in STRmix™): Hypothesis on the denominator of the LR (does not include POI)

Deconvolution: the process of separating a DNA profile into genotype sets and their associated probability weightings

LR: Likelihood Ratio

EPG: Electropherogram

CE: Capillary electrophoresis

OL: Off-ladder allele

LSAE: Locus specific allele efficiency

## 4 STRmix™ versions

The current STRmix™ version in use in Forensic DNA Analysis is v2.8.0. This version was implemented on 13 May 2021.

STRmix™ v2.8.0 is backwards compatible with v2.6 and v2.7 and therefore it is acceptable for a case to contain a mix of v2.6, v2.7.0 and v2.8.0 deconvolutions and/or LRs. In the situation where a profile has been previously deconvoluted using STRmix™ v2.6 and/or v2.7.0 and subsequent reference samples are received, the LRs may be run in v2.8.0 using the initial deconvolutions.


## 5 Software considerations

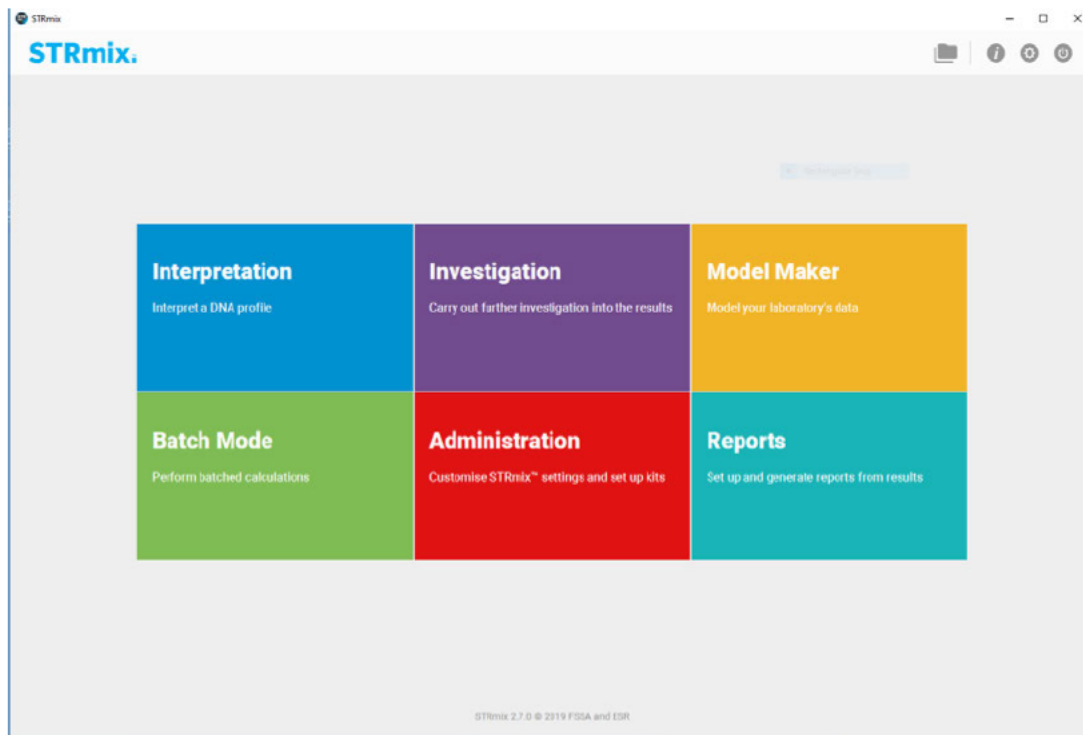
STRmix™ is stored on local computer drives.

STRmix™ cannot analyse DNA profiles that include text values such as OL or NR, nor can it analyse DNA profiles that contain mutations.

## 6 Using STRmix™ to perform a deconvolution

The main menu screen of STRmix™ is shown in Figure 1. The “Model Maker” and “Administration” functions are only to be used by high level users, e.g. scientists competent to train.

Clicking the  icon will open the results folder.



**Figure 1 – Main menu screen**

Clicking the “Interpretation” button produces the screen shown in Figure 2.

**Figure 2 – Interpretation screen**

The case number (e.g. QP19xxxxxxx) is entered in the “Case Number” field. The exhibit barcode number is entered in the “Sample ID” field. The “Case Notes” field should be used to enter the initials of the operator if the computer is logged in as ‘biology’, it can also be

used for adding further information such as assumptions about the analysis, but this is optional.

The “Number of Contributors” is typed in by the user. STRmix™ is validated for the analysis of DNA profiles with 1 – 4 contributors. STRmix™ requires that the scientist make a determination of the number of contributors to the DNA profile prior to performing the analysis. The number of contributors should be determined as per laboratory practice and is not within the scope of this document.

The “Run Settings” button will open the window shown in Figure 3. This should only be used to change the number of burn-in/post burn-in accepts, no other settings should be changed in this window. It should be noted that changing any of the “Run Settings” only affects the current deconvolution and does not affect global settings.

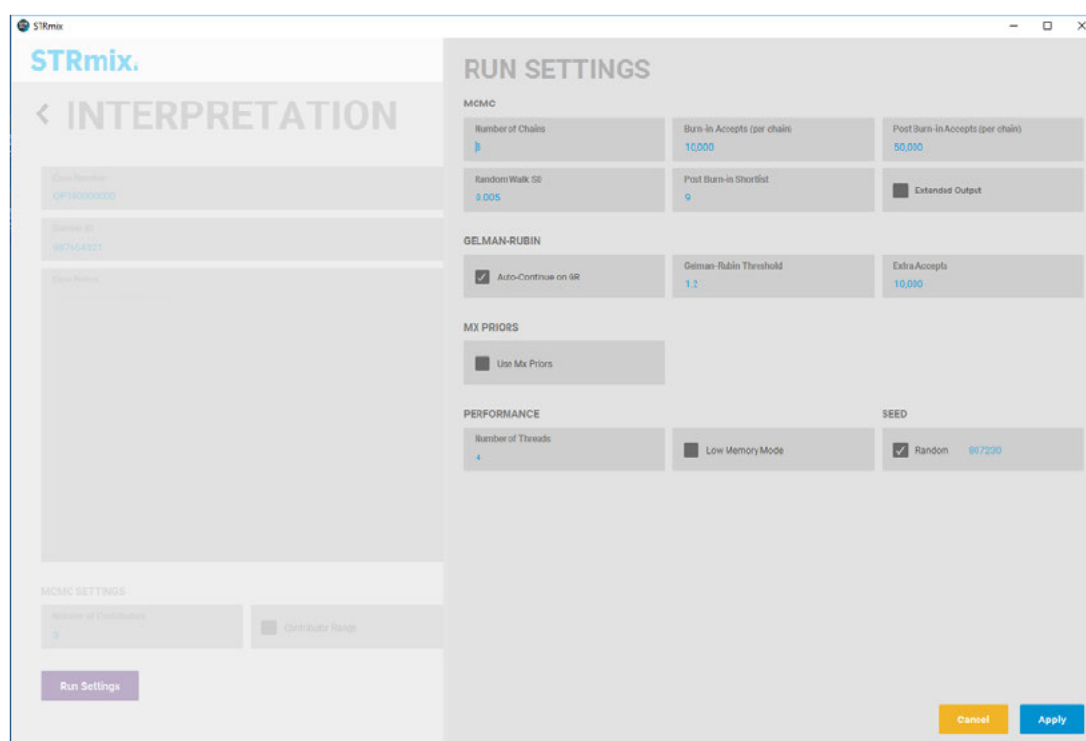


Figure 3 – Run settings screen

Once all information has been entered into the interpretation screen (Figure 2), click on the “Next” button. This will allow the operator to enter the input files on the screen shown in Figure { SEQ Figure \\* ARABIC }. The operator will select the profiling kit (either *PowerPlex\_21 3130* or *PowerPlex\_21 3500*) required for the profile being interpreted. Once the profiling kit has been selected the operator will enter casework input files either by drag and drop into the ‘Evidence Profile Data’ box or by navigating after clicking the ‘+’ button. Each input file that is required to be analysed for that sample should be entered into this box. Input files can be removed by highlighting the input file in question and clicking the “-” button. Reference sample input files are added to the ‘Reference Profile Data’ box in the same way as casework samples.

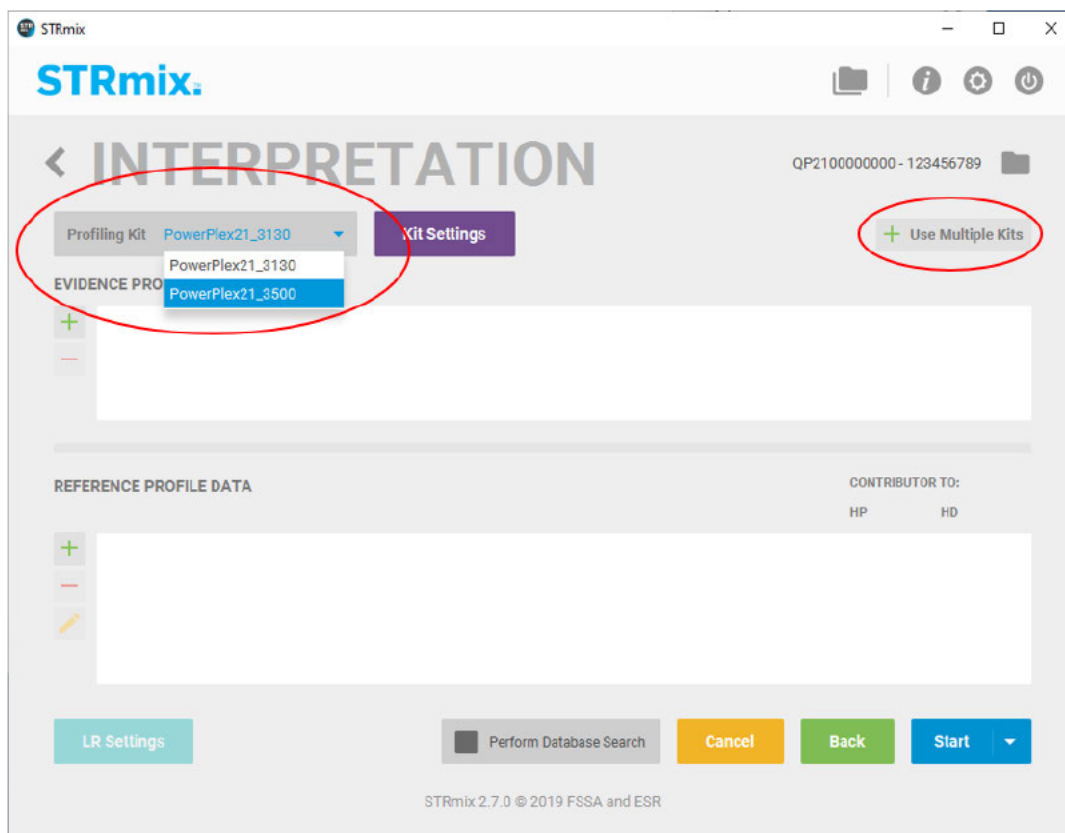


Figure { SEQ Figure \\* ARABIC } – Adding input files

If a casework sample has had multiple amplifications then all amplifications should be added to STRmix™ for analysis. The exceptions to this are when one of the profiles is excess and has 'OL' labels or if one of the amplifications appears to have an issue with it (for example, one of the amplifications is not representative of the other amplifications and therefore may not be modelled appropriately). If a sample is re-run (or re-CE'd) then only the re-run or the original run can be added to STRmix™, both cannot be added as they are from the same amplification.

If the casework sample has multiple amplifications processed on both the 3130xL and 3500xL Genetic Analysers then the combined kits functionality will be required. Click the 'Use Multiple Kits' button (Figure 4). This will prompt the operator to add the additional kit to the interpretation (this will differ depending on which kit is set as the default) as per Figure 4. Select the kit type for the first input file you wish to add (Figure 6 – Adding input files for multiple kits), then add the input file. Select the kit type for the next input file you need to add, then add the corresponding input file. Further information on adding input files for multiple kit analysis can be found in the most relevant STRmix Operation Manual.

For mixed DNA profiles, the 'Reference Profile Data' box should be left empty (unless performing a conditioned interpretation), for single source profiles the reference sample can be added at the same time as the casework sample if a LR is required to be calculated. For the comparison of reference samples see Section 7.

To perform a conditioned interpretation, the reference sample being conditioned on should be added to the 'Reference Profile Data' box and the HD box checked (see Figure 7).

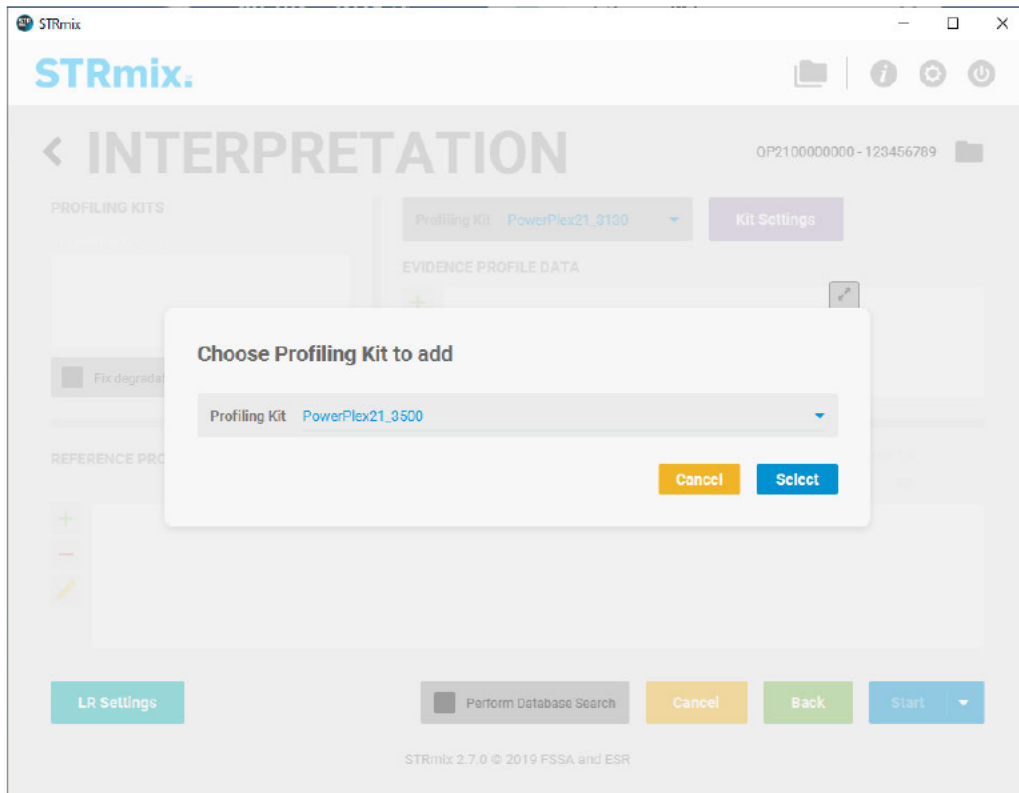


Figure 4 – Adding profiling kit

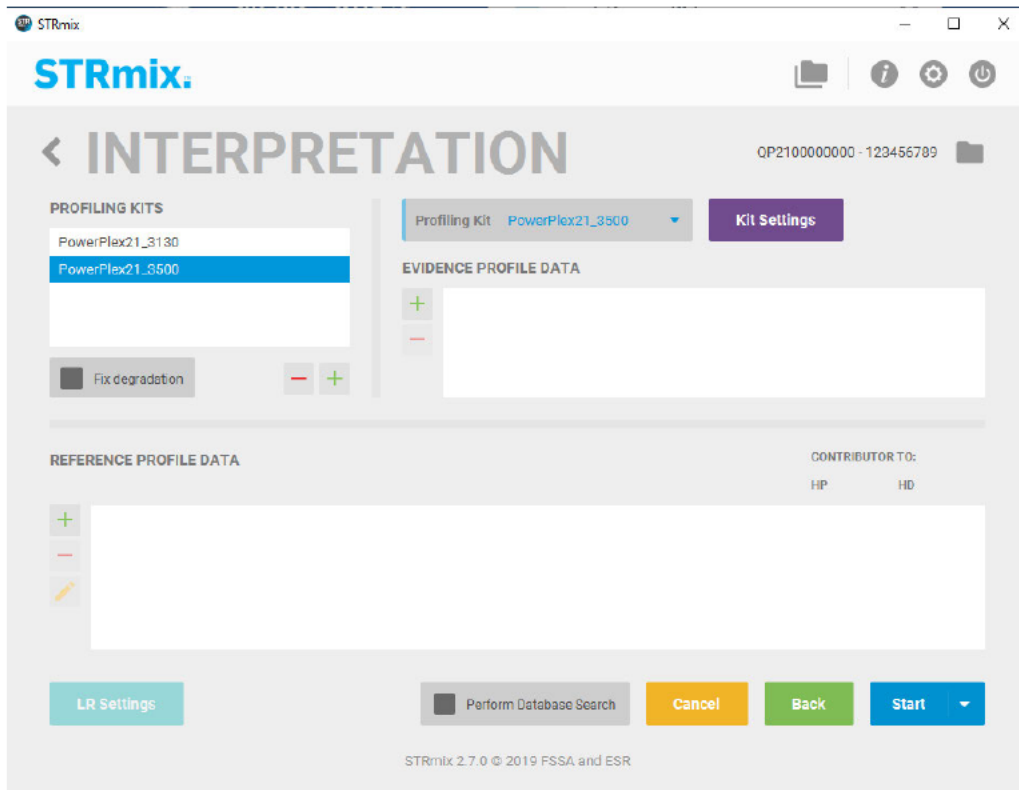


Figure { SEQ Figure \\* ARABIC } – Adding input files for multiple kits

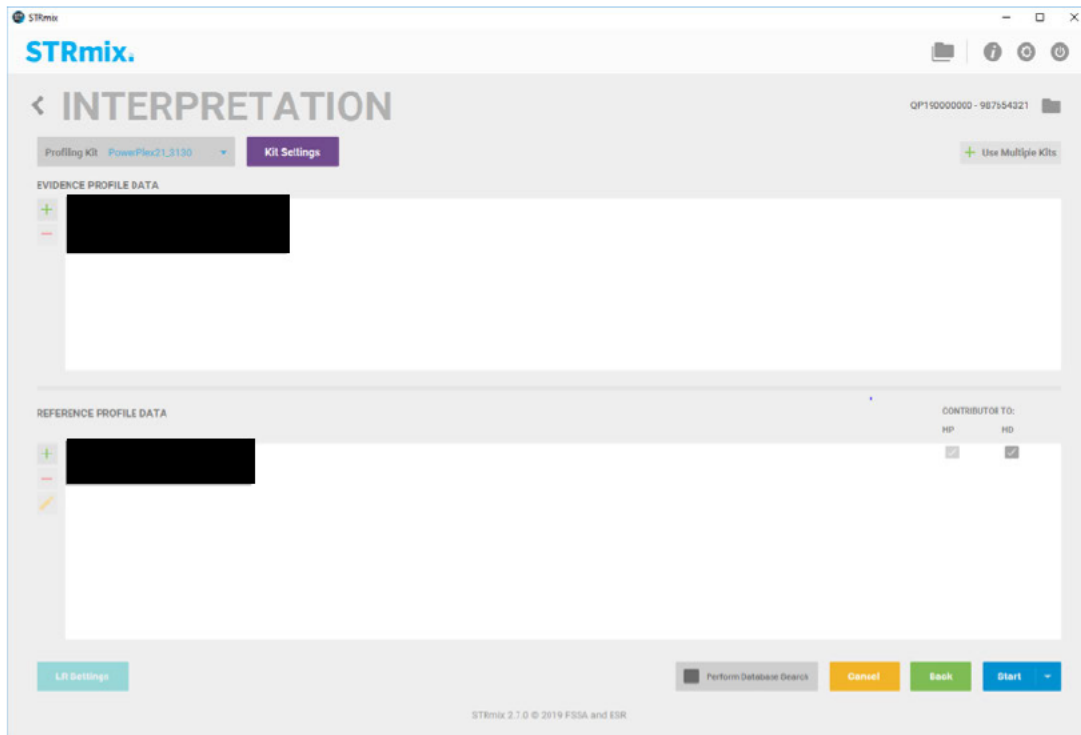


Figure { SEQ Figure \\* ARABIC } – Conditioning

Clicking the 'Kit Settings' button will open the screen shown in Figure 8.

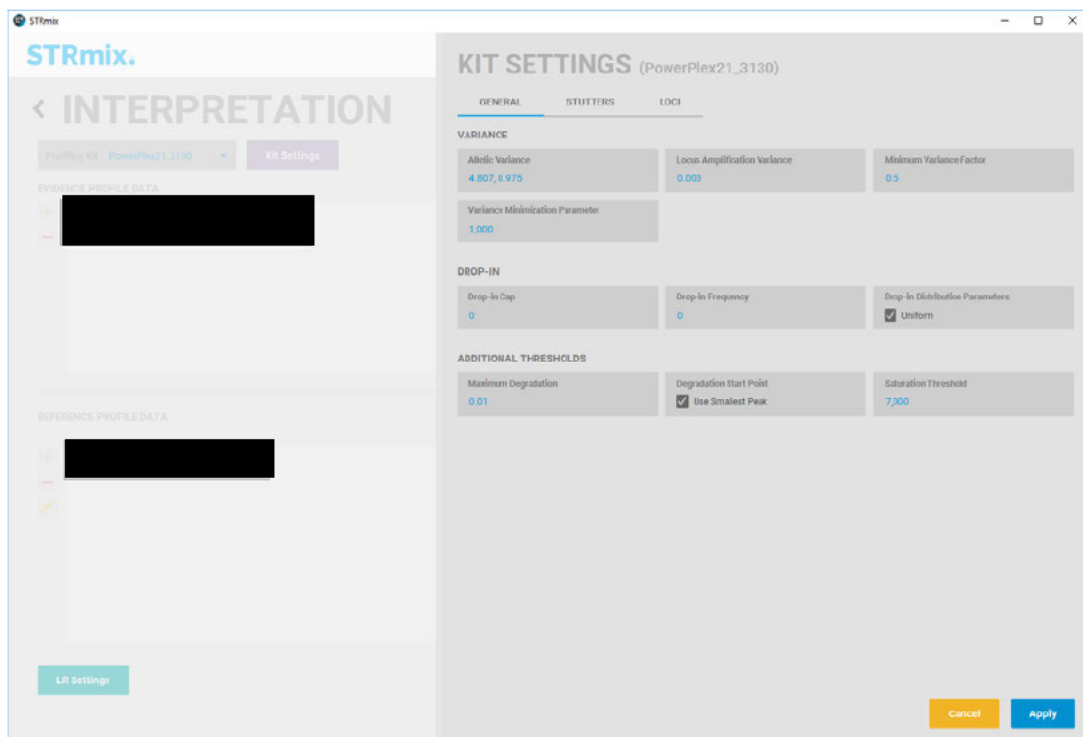


Figure 5 – Kit settings

Clicking on the 'Loci' tab will open the screen shown in Figure 9. From this screen, loci can be ignored, for example when a mutation is present in the profile.

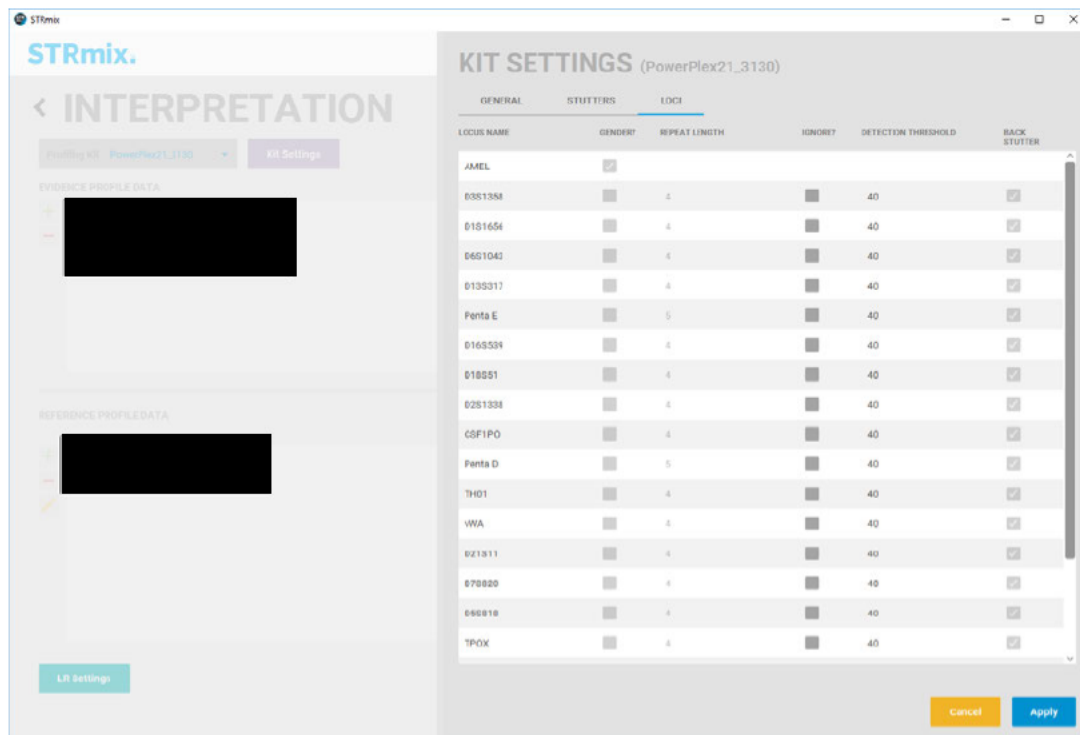


Figure 6 – Ignore loci

It should be noted that changing any of the “Kit Settings” only affects the current deconvolution and does not affect global settings.

Once the input files have been added to STRmix™, clicking the ‘Start’ button will start the analysis. STRmix will perform a stutter check and if any anticipated stutter peaks are missing, STRmix™ will display the following prompt (Figure 10).

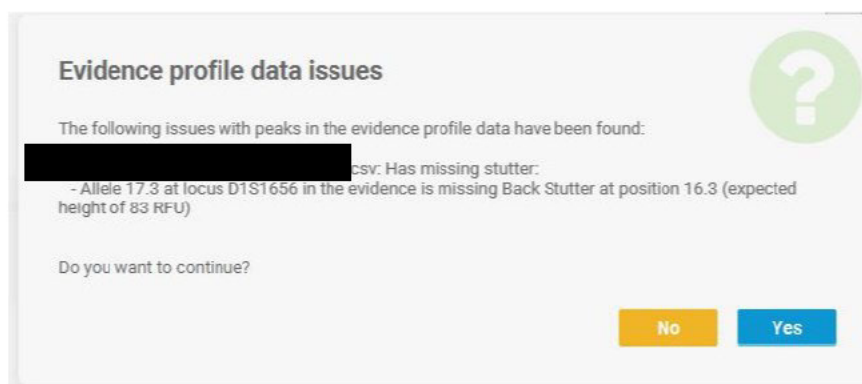


Figure 7 – Stutter check

If this prompt appears, check the profile to ensure the flagged stutter peak has not been unlabelled and/or investigate further. If the analyst is happy and wishes to proceed with the interpretation, click ‘Yes’. Otherwise click ‘No’ to cancel the setup.

The next screen (Figure 11) shows the progression of the analysis.

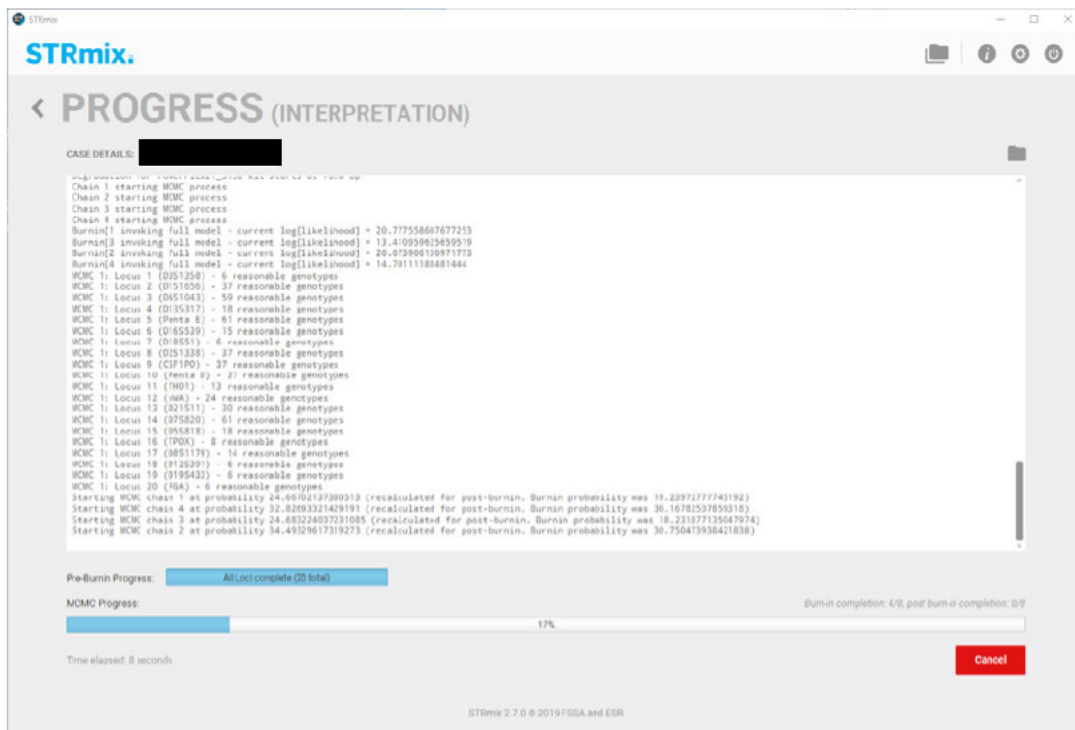


Figure 8 – Analysis progression

Once the analysis is complete, the screen shown in Figure 12 will display. Clicking the 'Finish' button will return the user to the main menu screen.

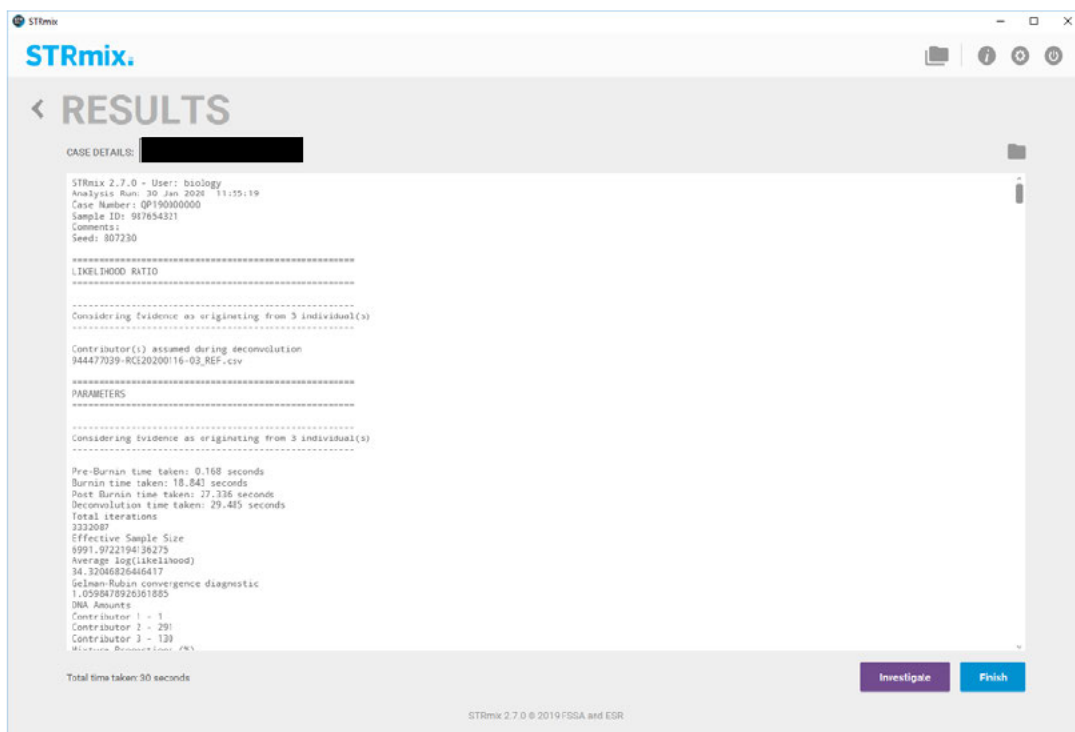


Figure 9 – Finish screen

Note: Some deconvolutions, especially those with multiple input files or those with more than two contributors, can take several hours to complete. If a deconvolution is to be left for



several hours, or overnight, it is important that the user does not log-off or shut down the computer.

Upon completion of the deconvolution the MoveSTRmixPDF Macro (QIS 35009) should be run. The macro will create a copy of the STRmix™ report and save it to the 'FR Upload' folder on the desktop. The results folder will be transferred to I:\STRmix Results.

## 7 Comparing reference samples to a deconvolution

With the exception of single source profiles and conditioned mixtures, all deconvolutions should be completed without reference samples. This is to ensure that all reference samples are compared to the same deconvolution therefore enabling LR from different reference samples to be directly compared with each other. LRs do not need to be calculated for reference samples that can be excluded from a DNA profile intuitively.

It is laboratory practice that reference samples are compared to DNA profiles separately as follows:

$$LR = \frac{\text{POI} + (n-1) \text{ unknowns}}{n \text{ unknowns}}$$

All LRs should be performed using the "Batch Mode" function to ensure the correct naming of the file for upload to the Forensic Register.

There are two ways to access the 'LR Batch' screen:

1. By clicking the 'Batch Mode' button on the main menu screen (Figure 1)
2. By clicking the 'Investigation' button on the main menu screen followed by 'Investigation Batch'

Clicking the 'Investigation' button will open the screen shown in Figure 13.

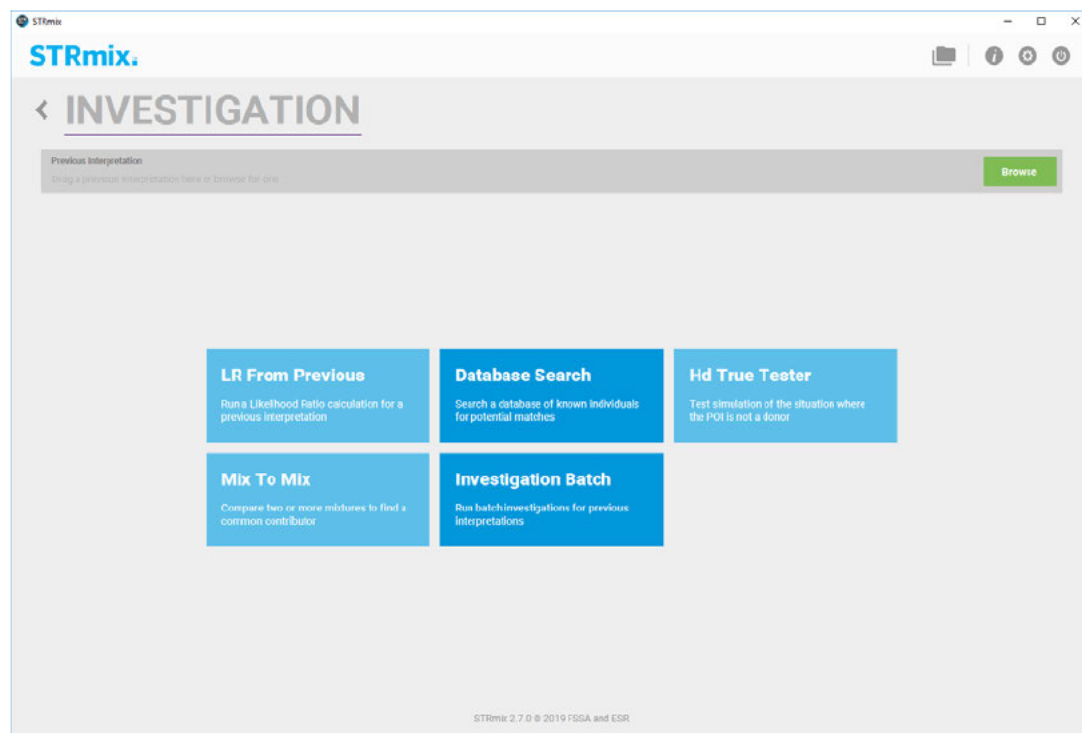


Figure 10 – Investigation screen

Clicking the 'Investigation Batch' button will open the screen shown in Figure 14. Drag and drop the entire deconvolution folder into the 'Interpretations' box. Alternatively, clicking the '+' button will allow the user to browse for the deconvolution file.

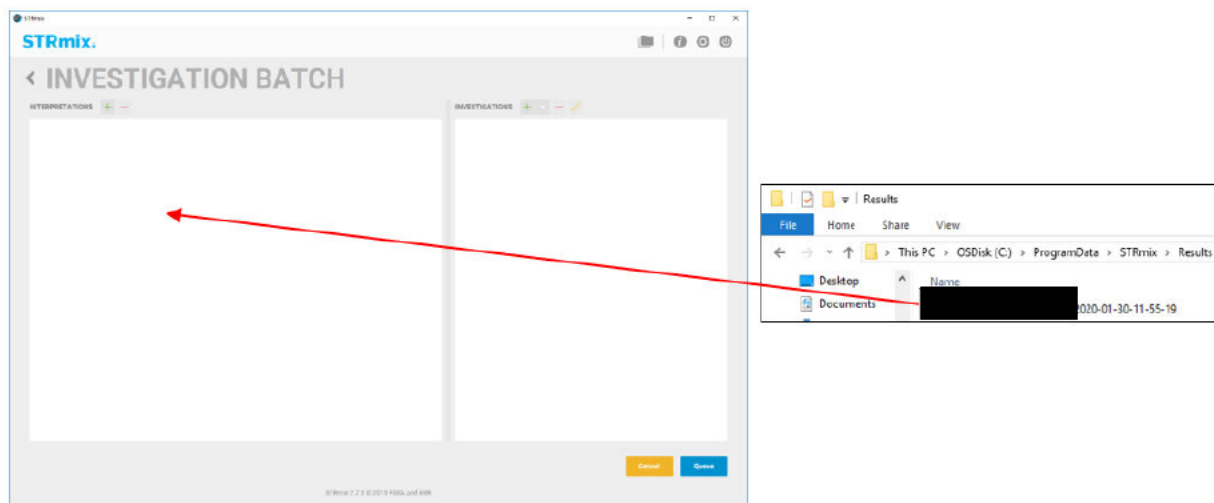



Figure 11 – Investigation Batch

Reference sample input files can be added by clicking either icons  above the 'Investigations' box. The screen shown in Figure 15 will open. The reference profile(s) for comparison can then be added by either drag and drop into the 'Reference Profiles' box or by pressing the + and locating the file.

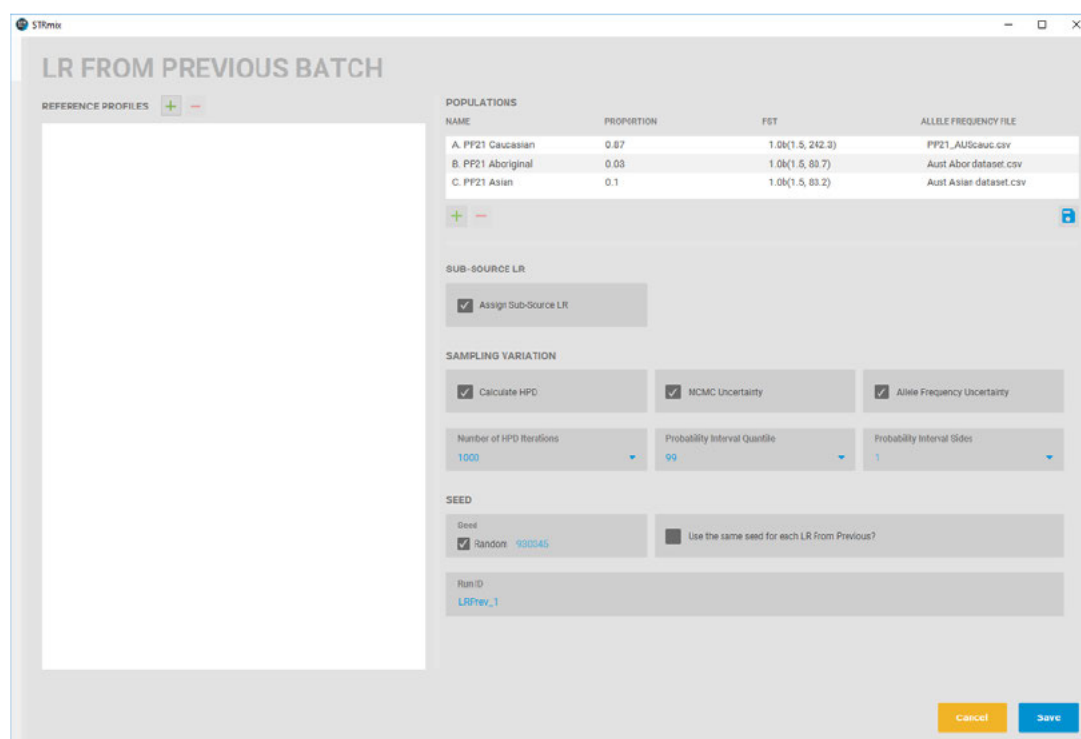
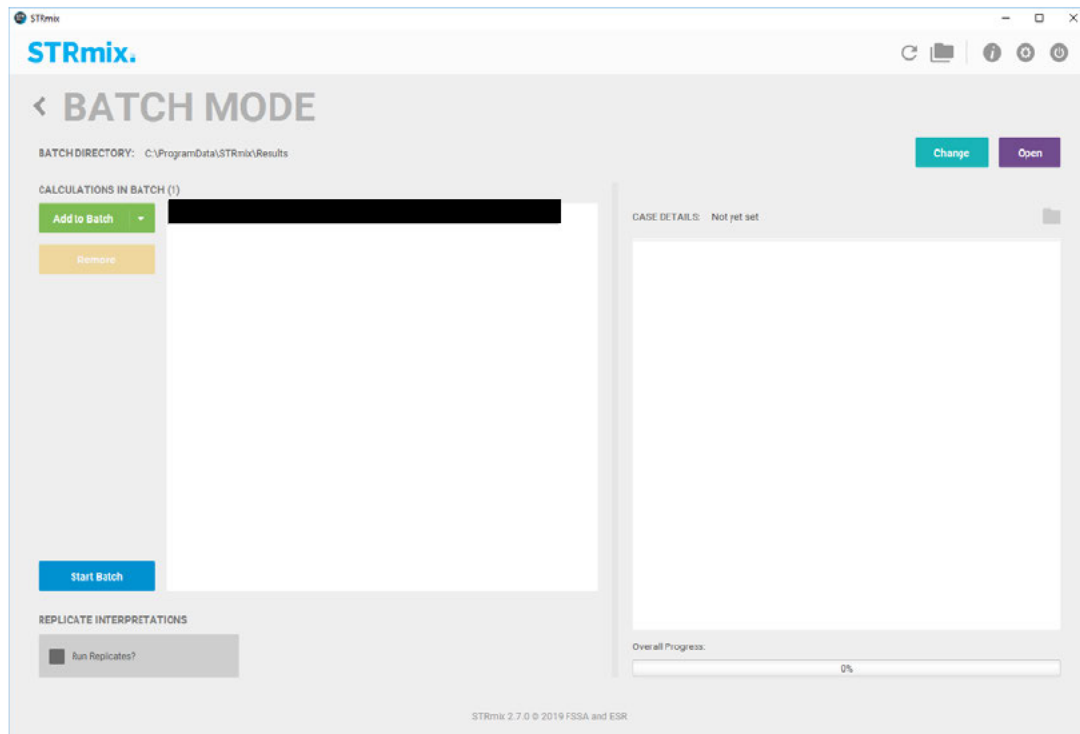


Figure 12 – Reference Profile Addition

Click the 'Save' button once all reference profiles have been added. This screen (Figure 15) will close and the previous screen (Figure 11) will be displayed. Click 'Queue' and the batch screen (Figure 16) will be displayed.



**Figure 13 – Batch screen**

Clicking 'Start Batch' will commence the LR batch processing.

## 8 Batch Mode

Clicking the 'Batch Mode' button from the Main Menu (Figure 1) will open the Batch Mode screen (Figure { SEQ Figure \\* ARABIC }).

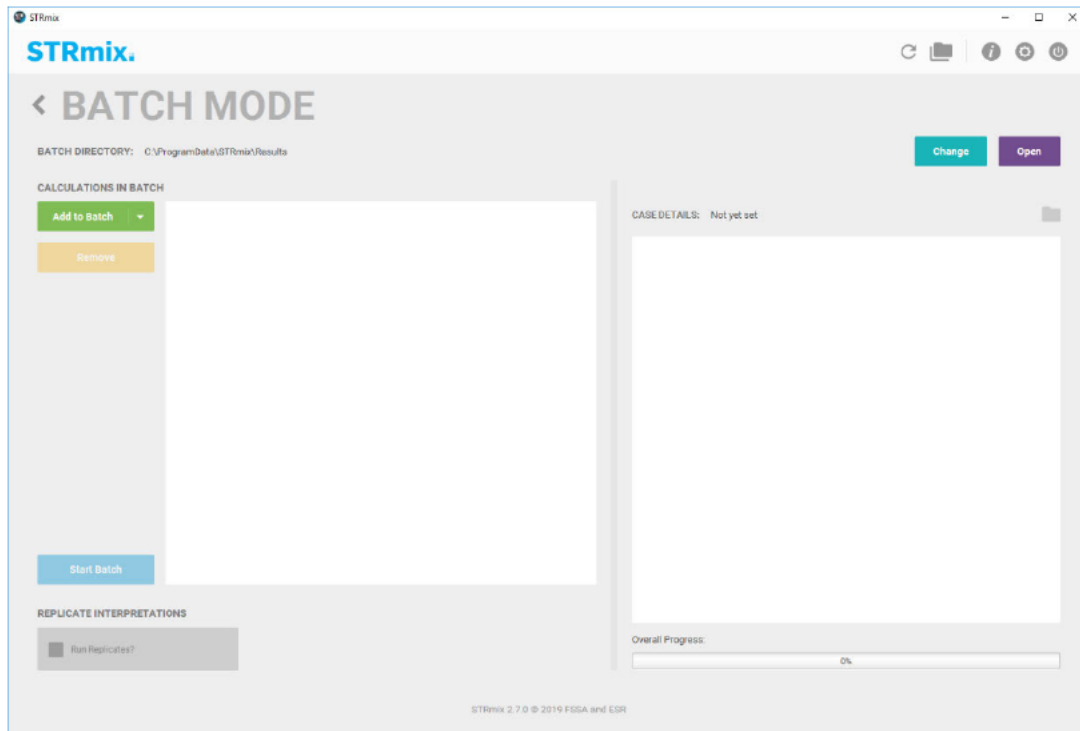


Figure { SEQ Figure \\* ARABIC } – Batch mode screen

From this screen deconvolutions and LR calculations can be added to a batch using the drop-down menu shown in Figure 18.

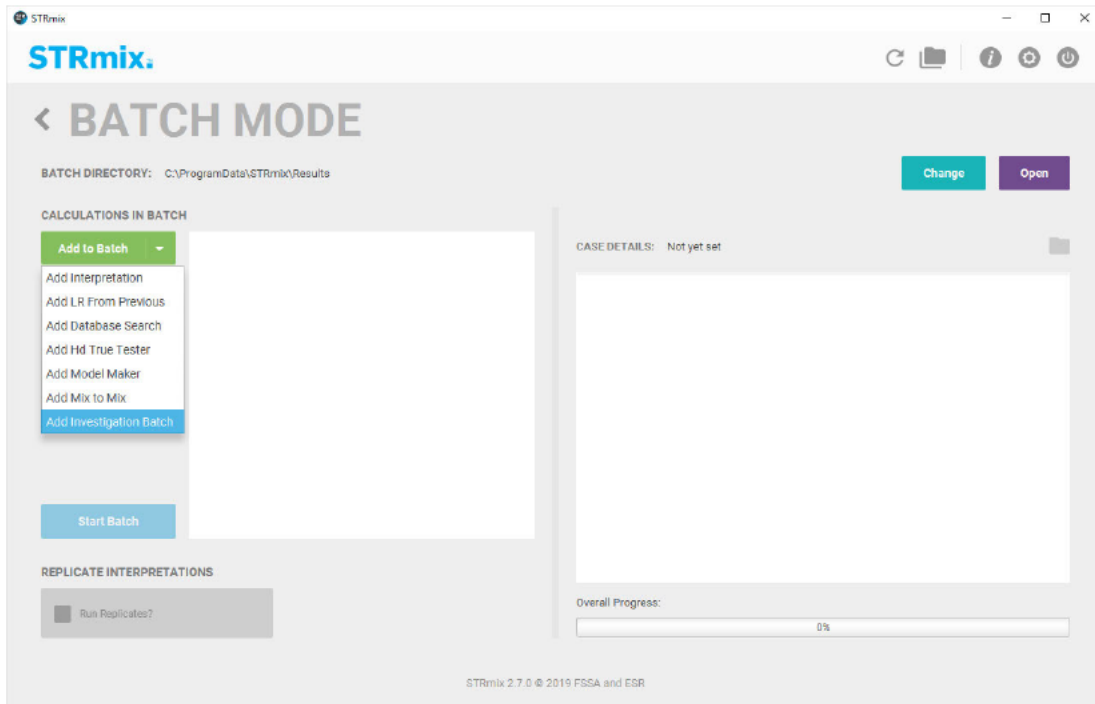


Figure 14 – Add to batch

Clicking on one of these options will open the interpretation screen (Figure 2) or the investigation batch screen (Figure 11) where analyses can be set up as per Sections 6 and 7.

## 9 Using the “Database Search” function

The Database Search function can be found by clicking the ‘Investigation’ button on the main menu screen.

The Database Search function has not been validated for use at this time.

## 10 Settings

STRmix™ uses laboratory specific parameters (determined from empirical data) in order to interpret a DNA profile. These include:

- Variance (calculated by Model Maker) – characterises the variation in peak heights within a DNA profile
- Detection threshold – the Limit of Reporting (LOR)
- Drop-in cap – the maximum height of a peak that can be considered to be drop-in
- Drop-in frequency – the frequency of occurrence of drop-in
- Drop-in parameters – the parameters that describe the Gamma distribution to which the laboratory drop-in data relates
- Saturation – the level above which peak heights are no longer proportional to the amount of DNA present in the sample due to saturation of the camera in the Genetic Analyser

Forensic DNA Analysis uses population proportions obtained from 2011 Census data to stratify the LR across the Australian Caucasian, Aboriginal and Asian population datasets. Theta ( $F_{st}$ ) is modelled from a distribution.

Forensic DNA Analysis quotes LRs at the sub-source level (using Factor of N!) using a 99% one-tailed highest posterior density (HPD) distribution.

The default settings used STRmix™ v2.7.0 and v2.8.0 in Forensic DNA Analysis are outlined in the Appendix.

## 11 Diagnostic tools

The diagnostic tools provide information about the performance of the MCMC including whether the MCMC has converged and how well STRmix™ has been able to explain the observed data.

The diagnostic tools should be checked with reference to the following guidelines:

Deconvolutions with a Gelman-Rubin convergence diagnostic (GR) value of  $>1.2$  (after additional iterations) should only be accepted for reporting after close scrutiny where the deconvolution is intuitively correct, and all other Diagnostic tools have given satisfactory results. Considerations should first be given to re-working and / or repeat STRmix™ analysis.

LSAE, Stutter and Allele variance values obtained at the point where the prior reaches its horizontal asymptote should only be accepted for reporting after close scrutiny where the deconvolution is intuitively correct, and all other Diagnostic tools have given satisfactory results. Considerations should first be given to re-working and / or repeat STRmix™ analysis.

Average Log(likelihood) values which are low or negative should only be accepted for reporting after close scrutiny where the deconvolution is intuitively correct, and all other Diagnostic tools have given satisfactory results. Considerations should first be given to re-working and / or repeat STRmix™ analysis.

Low Acceptance rate and Low Effective Sample Size (ESS) values may also indicate that the deconvolution has not performed optimally, and as such, should only be accepted for reporting after close scrutiny where the deconvolution is intuitively correct, and all other Diagnostic tools have given satisfactory results. Considerations should first be given to re-working and / or repeat STRmix™ analysis.

## 12 Troubleshooting

STRmix™ v2.7.0 has a known issue termed the 'Integer Overflow Issue' see STRmix, 2020, *Important STRmix Technical Update - Integer Overflow Issue*. [email] located in I:\STRmix settings\Manuals and Release Notes\STRmix v2.7. If the total number of iterations exceeds  $\sim 2.15 \times 10^9$  this could mean that the genotype weightings may be inaccurate. If this occurs solutions include reducing the number of amplifications added to the analysis. This issue has been resolved in v2.8.0.

If the analyst is expecting a profile to be resolved by STRmix™ to  $\geq 99\%$  for the purposes of upload to NCIDD and this does not occur, the analysis should be repeated with the number of burnin and post-burnin accepts doubled.

Refer to the most recent STRmix™ User's Manual for more STRmix™ troubleshooting.

## 13 Associated Documentation

QIS: 17117 – Procedure for Case Management

QIS: [33773](#) – Procedure for Profile Data Analysis using the Forensic Register

QIS: 35009 – MoveSTRmixPDF Macro

## 14 Validation

STRmix™ has been validated for use by the providers.

Forensic DNA Analysis has performed internal verification of STRmix™ v2.7.0 as Proposal #214 – Validation of STRmix™ v2.7.

Internal verification of STRmix™ v2.7.0 with profiles generated on the 3500xL Genetic Analyser has been performed under Proposal #219 – Verification of STRmix v2.7 for 3500xL.

Forensic DNA Analysis has performed internal verification of STRmix™ v2.8.0 as Proposal #231 – Validation of STRmix™ v2.8.

## 15 Records

STRmix™ User's Manuals and Operation Manuals can be found in I:\STRMix settings\Manuals and Release Notes.

## 16 References

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

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

J. M. Curran (2008) A MCMC Method for Resolving Two Person Mixtures. Science & Justice, 48, 168-177.

STRmix™ User's Manual

## 17 Amendment History

Version	Date	Updated By	Amendments
1	January 2019	E Caunt	First Issue
2	March 2020	H Pattison	Updated references and images from STRmix v2.6 to v2.7.0. Added new stutter check function.
3	March 2021	E Caunt	Removed reference to changing saturation setting, added multiple kits analysis, added 3500xL kit settings, reference to 'integer overflow issue' and doubling accepts if profile not resolved added
4	July 2021	E Caunt	Removed references to v2.7.0 to make more generic where possible. Updated Section 4. Updated database search to say not validated at this time. Added integer overflow issue resolved in v2.8.0 Updated template. Blank page removed

## 18 Appendix

### 18.1 Settings for profiles analysed on the 3130xI

#### MCMC Settings

Number of contributors	1
Use Mx priors	N
Number of chains	8
Burn-in accepts per chain	10,000
Post burn-in accepts per chain	50,000
Random walk SD	0.005
Post burn-in shortlist	9.0
Auto-continue on Gelman-Rubin	Y
Auto-continue on Gelman-Rubin threshold	1.2
Auto-continue on Gelman-Rubin accepts	10000

#### Kit Settings

Detection thresholds	D3S1358	40
	D1S1656	40
	D6S1043	40
	D13S317	40
	Penta E	40
	D16S539	40
	D18S51	40
	D2S1338	40
	CSF1PO	40
	Penta D	40
	TH01	40
	vWA	40
	D21S11	40
	D7S820	40
	D5S818	40
	TPOX	40
	D8S1179	40
	D12S391	40
	D19S433	40
	FGA	40
Saturation	7,000	
Degradation starts at	-1.0	
Degradation max	0.01	
Drop-in cap	0	
Drop-in rate parameter	0.0	
Drop-in parameters ( $\alpha$ , $\beta$ )	0.0, 0.0	
Min variance factor	0.5	
Variance minimization parameter	1,000	
LSAE variance parameter ( $1/\lambda$ )	0.003	
Allelic variance parameters ( $\alpha$ , $\beta$ )	4.807, 0.975	
<i>Back Stutter</i>		
Maximum stutter ratio	0.3	
Stutter variance parameters ( $\alpha$ , $\beta$ )	3.471, 1.425	



## Performance Settings

Number of threads	4
Low memory mode	N

## Kit Details

Size Regression File	PowerPlex21_SizeRegression.csv
<b>Back Stutter</b>	
Position Relative to Parent	-1, 0
Inversely Proportional To	Observed Height of Parent Allele
Stutter Regression File	PP21 Stutters.ini
Stutter Exceptions File	PP21 stutter exceptions_National.csv

## LR Settings

Number of populations	3
Assign sub-source LR	Y
HPD Iterations	1,000
Use MCMC uncertainty	Y
Use Allele Frequency uncertainty	Y
HPD quantile	99%
HPD sides	1
<b>A. PP21 CAUCASIAN</b>	
Proportion	0.87
FST	1.0b(1.5, 242.3)
Allele frequency file	PP21_AUScauc.csv
<b>B. PP21 ABORIGINAL</b>	
Proportion	0.03
FST	1.0b(1.5, 80.7)
Allele frequency file	Aust Abor dataset.csv
<b>C. PP21 ASIAN</b>	
Proportion	0.1
FST	1.0b(1.5, 83.2)
Allele frequency file	Aust Asian dataset.csv

## 18.2 Settings for profiles analysed on the 3500xL

**MCMC Settings**

Use Mx priors	N
Number of chains	8
Burn-in accepts per chain	10,000
Post burn-in accepts per chain	50,000
Random walk SD	0.005
Post burn-in shortlist	9.0
Auto-continue on Gelman-Rubin	Y
Auto-continue on Gelman-Rubin threshold	1.2
Auto-continue on Gelman-Rubin accepts	10000

**Kit Settings**

## KIT SETTINGS

KIT SETTINGS		
Ignored loci		
Detection thresholds	D3S1358	80
	D1S1656	80
	D6S1043	80
	D13S317	80
	Penta E	80
	D16S539	80
	D18S51	80
	D2S1338	80
	CSF1PO	80
	Penta D	80
	TH01	80
	vWA	80
	D21S11	80
	D7S820	80
	D5S818	80
	TPOX	80
	D8S1179	80
	D12S391	80
	D19S433	80
	FGA	80
Saturation	30,000	
Degradation starts at	-1.0	
Degradation max	0.01	
Drop-in cap	250	
Drop-in rate parameter	0.0264	
Drop-in parameters ( $\alpha$ , $\beta$ )	0.31, 199.83	
Min variance factor	0.5	
Variance minimization parameter	1,000	
LSAE variance parameter ( $1/\lambda$ )	0.01	
Allelic variance parameters ( $\alpha$ , $\beta$ )	10.197, 1.801	
<b>Back Stutter</b>		
Maximum stutter ratio	0.3	
Stutter variance parameters ( $\alpha$ , $\beta$ )	1.703, 14.134	
<b>+1 rpt stutter</b>		
Maximum stutter ratio	0.1	
Stutter variance parameters ( $\alpha$ , $\beta$ )	5.519, 28.11	

## Performance Settings

Number of threads	4
Low memory mode	N

## Kit Details

Size Regression File	PowerPlex21_SizeRegression.csv
<i>Back Stutter</i>	
Position Relative to Parent	-1, 0
Inversely Proportional To	Observed Height of Parent Allele
Stutter Regression File	PP21 Stutters.ini
Stutter Exceptions File	PP21 stutter exceptions_National.csv
<i>+1 rpt stutter</i>	
Position Relative to Parent	1, 0
Inversely Proportional To	Observed Height of Parent Allele
Stutter Regression File	PowerPlex21_ForwardStutter.txt
Stutter Exceptions File	

## LR Settings

Number of populations	3
Assign sub-source LR	Y
Calculate HPD	Y
HPD Iterations	1,000
Use MCMC uncertainty	Y
Use Allele Frequency uncertainty	Y
HPD quantile	99%
HPD sides	1
<i>A. PP21 CAUCASIAN</i>	
Proportion	0.87
FST	1.0b(1.5, 242.3)
Allele frequency file	PP21_AUScauc.csv
<i>B. PP21 ABORIGINAL</i>	
Proportion	0.03
FST	1.0b(1.5, 80.7)
Allele frequency file	Aust Abor dataset.csv
<i>C. PP21 ASIAN</i>	
Proportion	0.1
FST	1.0b(1.5, 83.2)
Allele frequency file	Aust Asian dataset.csv
PERFORMANCE SETTINGS	
Number of threads	4
Low memory mode	N