2.8 Operation Manual

STRmix™ V2.8 Operation Manual

Issued by Institute of Environmental Science and Research Limited

Date of Issue: 29 September 2020

CONTENTS

1. INTRODUCTION

1.1. Background

The *STRmix™ V2.8 Operation Manual* provides information about and instructions for using the expert interpretation software STRmix™. It should be read in conjunction with the *STRmix™ V2.8 User's Manual*, which also contains a summary of useful nomenclature, abbreviations and an explanation of some mathematical terms that readers may find a useful reference.

Manual conventions: All references to on-screen items are capitalised, e.g. the Startup screen. On-screen items that are subject to an action are written in bold text, e.g. select **Start**.

1.2. An overview of STRmix™ V2.8 capabilities and limitations

STRmix™ V2.8 has the following capabilities:

1. STRmix™ can be used to interpret DNA profiling data generated using any autosomal STR profiling kit with information at more than one locus
2. Deconvolution of mixed DNA profiles containing any specified number of contributors (or range of contributors) (within limits of computing capacity) without reference to any known contributors or persons of interest (POI)
3. Comparison of reference DNA profiles to single source and mixed DNA profiles of any number of contributors (within limits of computing capacity) to generate a measure of weight of the evidence (likelihood ratio) in relation to a pair of propositions, including assessing scenarios involving familial relationships
4. Calculate a likelihood ratio by comparing reference DNA profiles from POIs to a genotype probability distribution produced by a previous STRmix™ analysis
5. Identify potential contributors to mixed DNA profiles by comparing an output file from a previously deconvoluted profile against a database of known reference profiles, using either a direct or familial database search function
6. Compare two or more mixed DNA profiles and assign an LR to help address whether or not a common donor exists between any of the samples
7. Incorporation of multiple epgs (replicate amplifications generated using the same STR multiplex) into one analysis
8. Incorporation of multiple epgs generated via different protocols (e.g. instrument types) or STR kits into one analysis
9. Calculate multiple Likelihood Ratios using multiple reference samples for one deconvolution
10. Accommodate data generated via protocols demonstrating increased stochastic variation and non-zero allelic drop-in rates (e.g. elevated PCR cycle number and enhanced CE injection methods)
11. Perform a Top Down Approach to DNA mixture interpretation by determining an LR to the top k contributors to a mixed DNA profile

STRmix™ cannot:

1. Analyse lineage marker data and indels
2. Account for mutation events or triallelic loci within its calculations

There are eight parameters which are not optimised by the Markov chain Monte Carlo (MCMC) in a STRmix™ analysis, these are:

1. Number of contributors within a standard deconvolution
2. Observed peak heights and their molecular weights
3. Stutter ratios
4. Capillary electrophoresis camera saturation
5. Detection thresholds
6. Drop-in parameters
7. The locus specific amplification efficiency (LSAE, A^l) variance prior distributions
8. Allele and stutter peak height variance prior distributions.

These parameters must either be modelled or entered at analysis. The number (or range) of contributors is specified by the user at the start of an analysis and observed peak heights and molecular weights are obtained from the evidence input file. The hyper-parameter for standard deviation of A^l values and the peak height variances are values drawn by the model from analysis of profile data. Model Maker within STRmix™ is used to determine the optimal variance values to use in the above processes, based on data generated in your laboratory, using laboratory specific protocols and instrumentation platform(s) for each STR multiplex employed. The use of Model Maker is described later in this manual. The values used for stutter ratio files, drop-in parameters, detection thresholds and camera saturation should be taken from empirical data. Refer to the relevant sections within this manual for further detail.

2. FORMATTING INPUT DATA FOR STRMIX™

2.1 Allowable and rare alleles

STRmix™ requires input data describing an electropherogram (epg) in order to run. Like all software, STRmix™ requires data in a specific format for it to be entered. As STRmix™ models the height of both allelic and stutter peaks, stutter must not be removed from evidence samples at epg analysis. Labels must be removed from all other analysis artefacts including pull up, spikes and dye blobs.

Non numeric values such as OL, < or > and R are not permitted within the STRmix™ input files. Unambiguous alleles including those that are rare should appear in the corresponding input file as their actual allelic size designation, for example D21 30.1.

There is no function to accommodate somatic mutations or triallelic loci in STRmix™ calculations. If a profile has a triallelic STR pattern, STRmix™ can still use the input data if the locus is ignored (see section 3.2.2 Add evidence profile data).

2.2 Input file format

There are two types of epg summaries that can be input into STRmix™:

- Evidence (also referred to as crime or questioned) epg summaries (which require locus, peak designations, size - i.e. molecular weight - and height information).
    - As STRmix™ models stutter you will need to turn off any stutter filter during analysis to capture potential stutter information within your input file. Within V2.8, you can model any type of stutter, including (but not limited to) expected forward stutter (plus one full repeat), expected back stutter (minus one full repeat), expected double back stutter (minus two full repeats) and expected 2 base pair stutter data.
- Reference (also referred to as knowns or standards) epg summaries (which require locus and allelic designations).
    - As STRmix™ does not *interpret* the reference profile, do not include stutter data in your reference input file.

There are two types of input file format; text files (extension .txt, referred to as text files in this document) and comma separated values files (extension .csv, referred to as STRmix™ files). Text files are created using software used to analyse CE data. For an evidence input file from GeneMapper™ a genotype table is required. To set up the appropriate table template in GeneMapper™ undertake the following steps:

Step	Action
1	To display the Table Settings tab, select Tools>GeneMapper™ manager>Table Settings. Select **New** to open the table Setting Editor window
2	In the General tab, enter a suitable name for the table setting, e.g. STRmix™ casework.
3	In the Genotypes tab, select Sample Name, Marker, Allele, Size and Height. Sort by sample, then dye, then marker in the drop down menus.
4	Enter the appropriate number of alleles per marker.
5	Optionally, deselect ‘Keep Allele, Size, … and Comment together’ box

6	Select **OK** to save the table setting.
7	A template for reference input files can be created in the same manner except that size and height are not required.

Notes

If using OSIRIS for data analysis, contact the OSIRIS team for an appropriate export file.

If using GeneMarker® for data analysis, instructions are available from the STRmix™ Technical and Scientific Support website:

http://support.strmix.com/support/solutions/articles/1000235676-analysis-using-genemarker

Commas within sample names will be removed (“,”).

There is no requirement for reference and crime profiles to be generated using the same multiplex.

On the completion of analysis, select the appropriate STRmix™ genotype table setting that has been created as per the above instructions. To export the information, select **File** > **Export table** whilst in the Genotypes analysis tab of GeneMapper™.

Text files can be saved in the following order, either:

Sample name, Marker, Allele1, .., Allele n, size1, ..., size n, height1, ..., height n

Or

Sample name, Marker, Allele1,size1, height 1 .. Allele n, size n, height n

Where n is the actual number of peaks selected in the export file and is user configurable.

There is no requirement to duplicate homozygote alleles within the crime sample or reference text file. Alleles will be duplicated with the STRmix™ .csv reference file.

STRmix™ files may be created manually in Excel or Notepad or similar. Text files are converted to STRmix™ .csv files after successful import and are saved in the appropriate results folder. STRmix™ will convert any single homozygote designations within a reference .txt file to a duplicate designation within the .csv reference STRmix™ file. Examples of STRmix™ evidence and reference input files are given in Figure 2.1 and Figure 2.2. STRmix™ will append evidence files with _EV and reference files _REF.

If creating your own .csv STRmix™ file to input as a reference, then any homozygote alleles must be duplicated.

Leaving stutters on in a reference will cause an input file error.

	A	B	C	D
1	Locus	Allele	Height	Size
2	D8S1179	11	134	135.34
3	D8S1179	12	2106	139.56
4	D8S1179	14	25	147
5	D21S11	27	53	196.43
6	D21S11	28	1004	200.3
7	D21S11	29	1025	204.34
8	D7S820	9	47	267.59
9	D7S820	10	1272	271.59
10	CSF1PO	9	39	316.43
11	CSF1PO	10	934	320.43
12	CSF1PO	11	875	324.51
13	D3S1358	15	149	123.7
14	D3S1358	16	2191	127.76
15	TH01	6	26	171.08

Figure 2.1 A partial example of a STRmix™ evidence input file (.csv format)

	A	B
1	Locus	Allele
2	D8S1179	10
3	D8S1179	12
4	D21S11	30
5	D21S11	31.2
6	D7S820	11
7	D7S820	11
8	CSF1PO	10
9	CSF1PO	12
10	D3S1358	15
11	D3S1358	17
12	TH01	7
13	TH01	7

Figure 2.2 A partial example of a STRmix™ reference input file (.csv format)

3. RUNNING STRMIX™

3.1 Launching STRmix™

Start STRmix™ by double-clicking the STRmix™ icon on the desktop or select STRmix™ from **Start > STRmix™**. The first time that STRmix™ is run on a computer, the user will be prompted to set a password (see Figure 3.1).

Please note that this password is unable to be changed and will be required to be used for all changes made to the settings including changes to stutter files, kits, populations and allele frequencies as well as saving defaults (e.g. for Database search). If you do not wish to have an Administrator password, select **Don't Protect**. This will mean that the user will not be asked for a password again, and all future changes to settings can be made without a password.

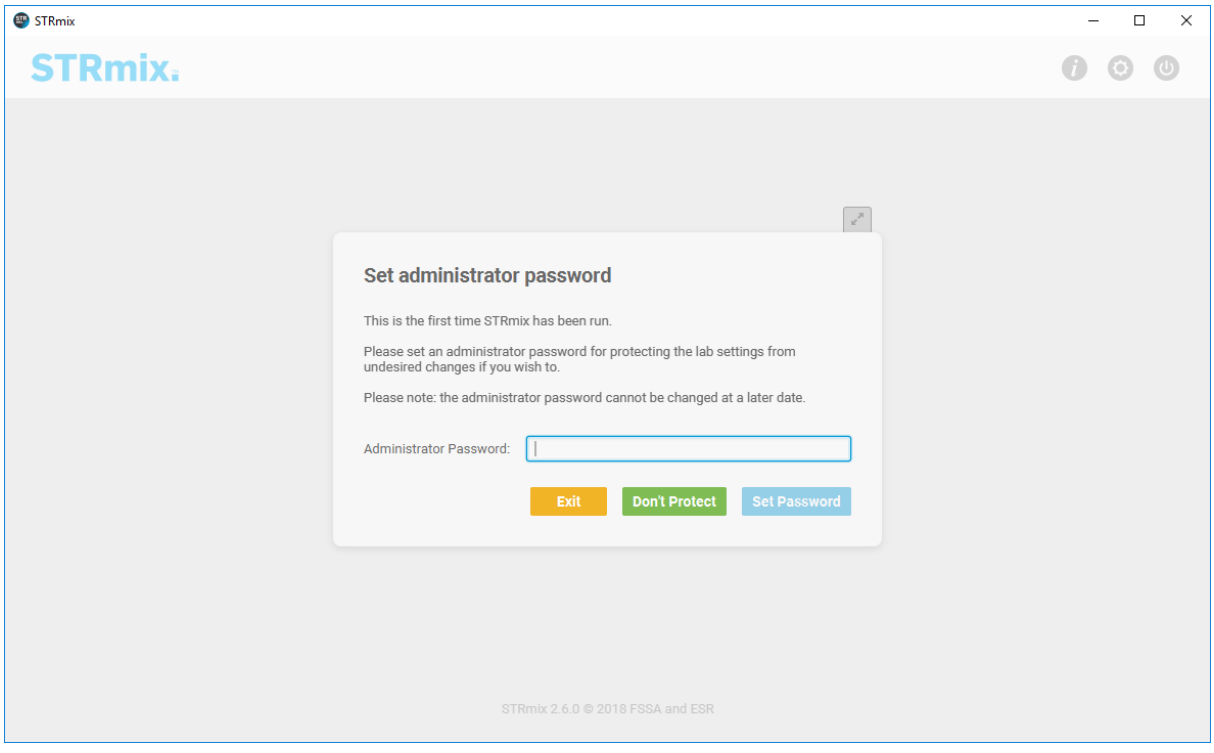


Figure 3.1 Password prompt on starting STRmix™ for the first time.

Once a password has been set (or this step has been bypassed), the STRmix™ Main Menu will display (see Figure 3.2).

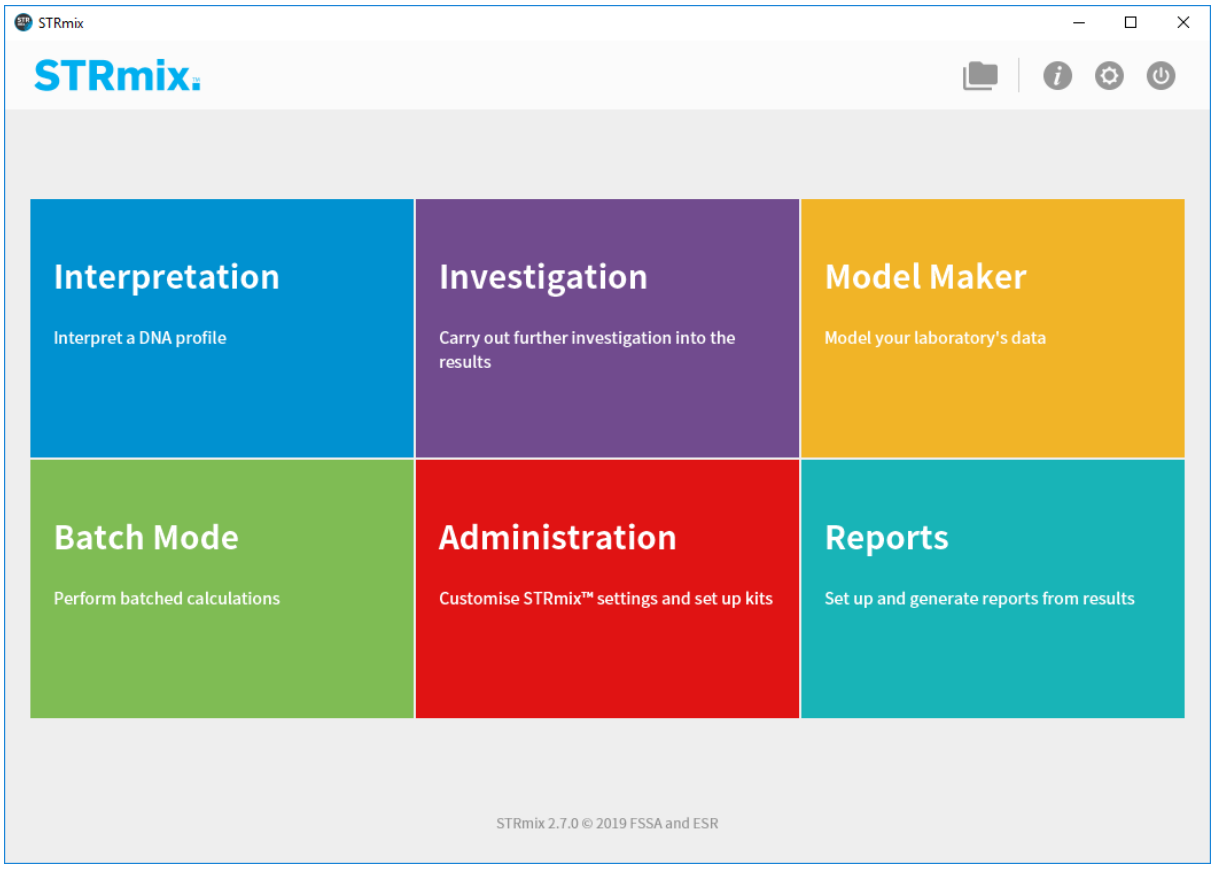


Figure 3.2 STRmix™ Main Menu

Each of the functions within STRmix™ can be accessed from the Main Menu and are discussed in turn within this manual. Selecting the light blue **STRmix™** logo (top left of window) at any time will return the user to the Main Menu.

3.2 STRmix™ Tool bar

The STRmix™ toolbar (see Figure 3.3) is available from all windows within STRmix™. The options are described in Table 3-1 STRmix™ toolbar options. Note that different options may be available from different windows.

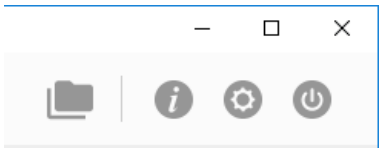


Figure 3.3 STRmix™ toolbar

Table 3-1 STRmix™ toolbar options

Symbol	
	Open results directory

	Open the About window. Includes system requirements, third-party software information, and software licence agreement. From this window, the licence can be removed
	Open the preferences: 1. Toggle the theme from light to dark 2. Allow the display language to be changed
	Exit STRmix™

For information on how to add custom languages, see support.strmix.com

3.3 Interpretation

Select Interpretation to start a deconvolution (see Figure 3.4). This initiates the process for an interpretation of a single source or mixed DNA profile.

3.3.1 Case details

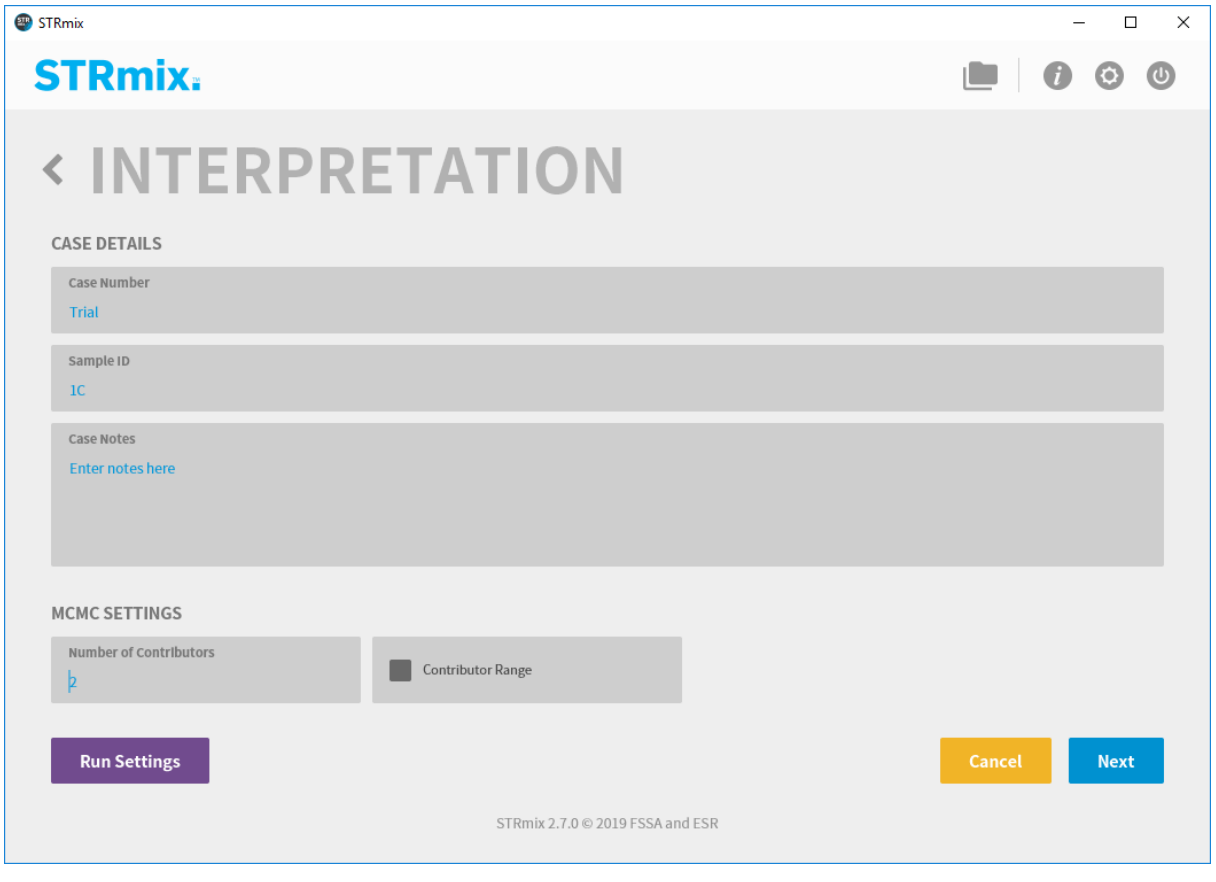
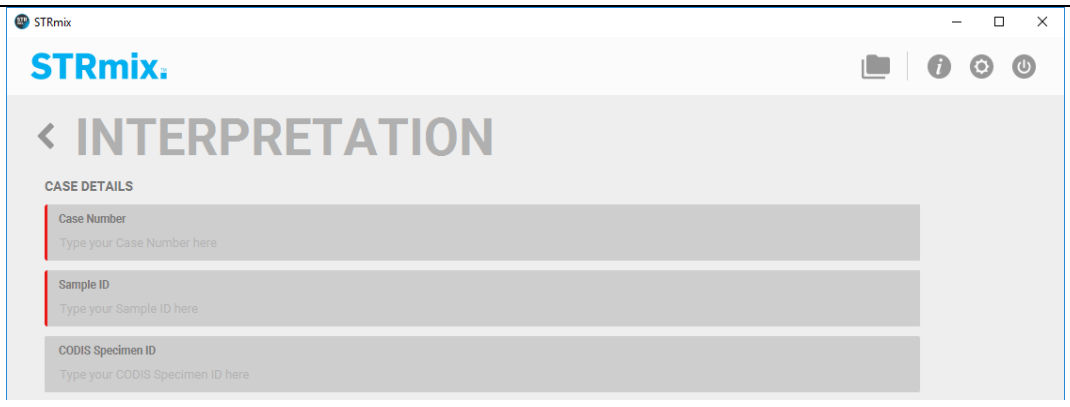
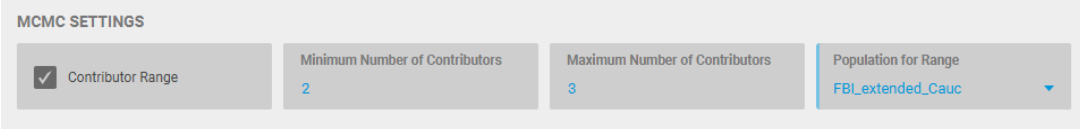


Figure 3.4 STRmix™ Interpretation case details window

Step	Action
1	Complete case number and sample ID information and optionally enter case notes. If CODIS reporting is enabled, the CODIS Specimen ID field will be enabled. This ID will be written to the CODIS xml report.



2 Update number of contributors. To interpret a range of contributors, select **Contributor range** and enter the range of contributors. Note that STRmix™ has only been developmentally validated for a contributor range difference no greater than one. Select a population for range from the dropdown box. A pragmatic choice is to align the Population for Range with the population database used in the *LR* that has the largest population proportion. This choice is not expected to have a significant difference on the *LR*. Please refer to the STRmix™ 2.8 User's Manual for more detail.



3 Select **Run Settings** (

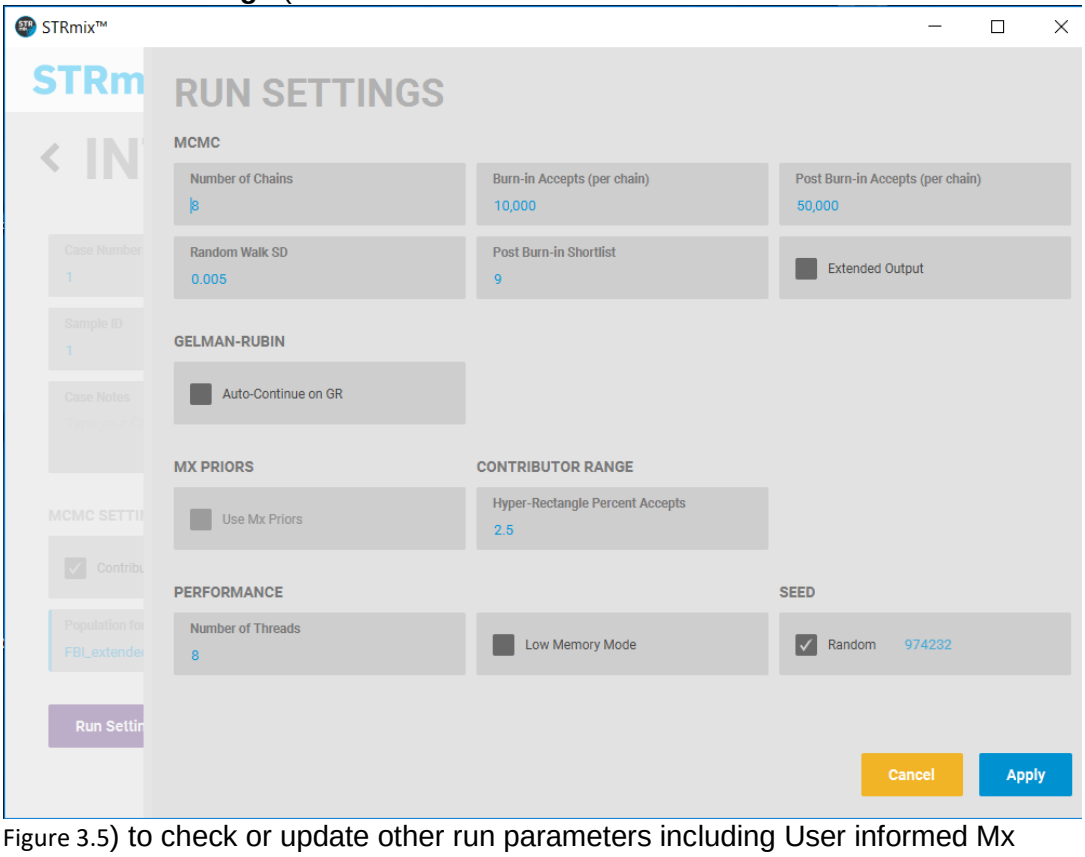


Figure 3.5) to check or update other run parameters including User informed Mx priors (refer section 3.3.8), if required and **Apply.**

4	Select **Next** to proceed to Add profile data (refer section 3.3.2) or **Cancel** to return to the Startup screen/Main Menu.

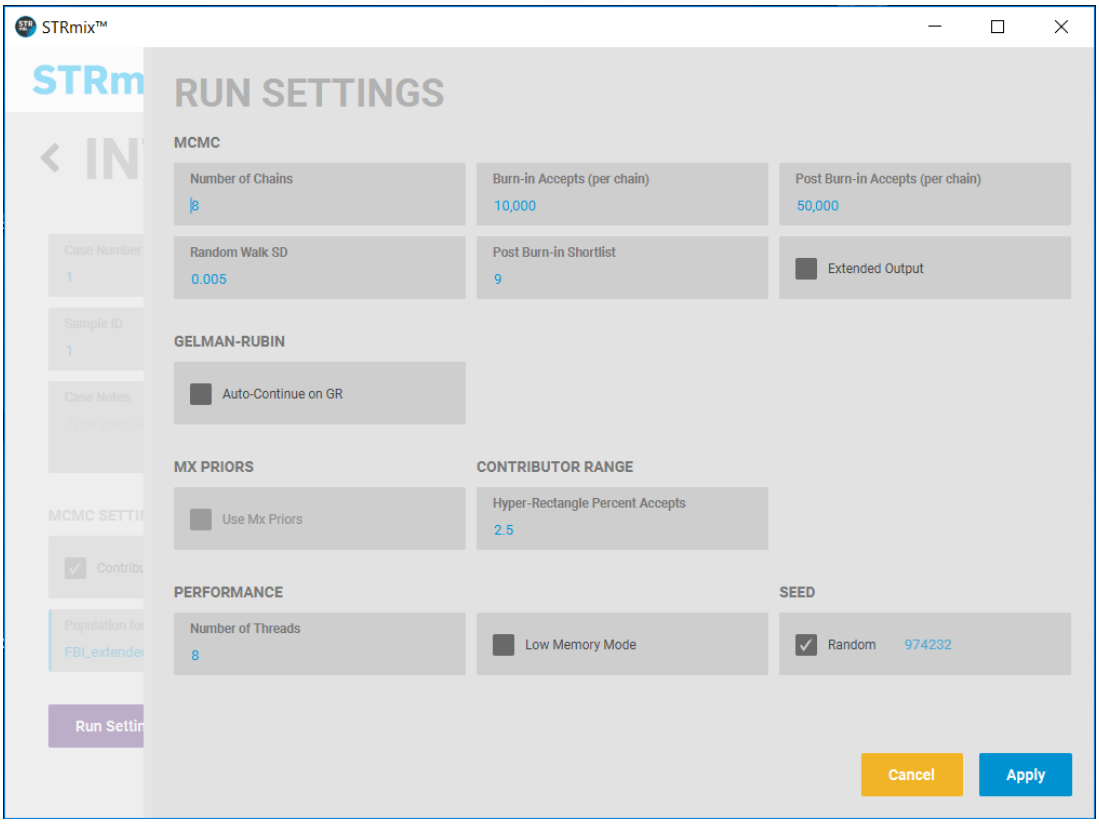


Figure 3.5 Run Settings

Notes

STRmix™ results folders are prefixed with the information entered into the Case Number field, followed by sample ID, then the date and time they were created in the format yyyy-mm-dd-hh-mm-ss.

The MCMC Settings are default parameters (refer section 7.1 Default settings). Settings may be changed from the Run Settings screen (see

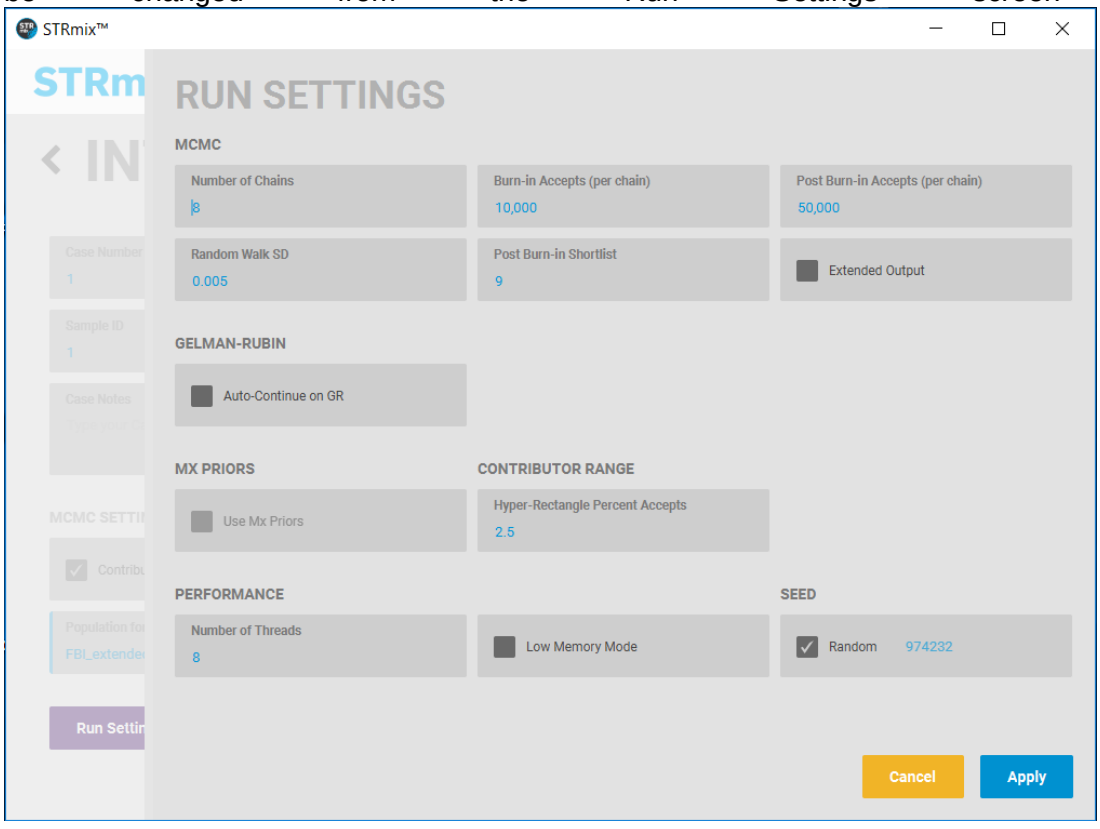


Figure 3.5) for individual sessions with no requirement for password input, but are not saved as the system default on run completion. The seed is the starting number used within the random number generator. Setting the seed to the same state will return the same results from run to run and is used to assist with validation.

The Number of Contributors should be checked prior to each run. The number of contributors is entered as free text in the appropriate box. There is no limit to the number of contributors STRmix™ can interpret however for practical purposes this may be limited by computing power. A range of contributors can be interpreted within STRmix™ V2.8.

The number of Burn-in Accepts (per chain) and number of Post Burn-in Accepts (per chain) defaults (10,000 and 50,000 respectively) have been selected as optimum values and will be suited to the majority of samples run. Assuming eight chains are run then the total number of Post Burn-in Accepts will be 400,000. The number of accepts may be decreased for single source profiles where there is no ambiguity in the profile and they may be increased for complicated mixed and/or sub-optimal DNA profiles such as those that are inhibited, degraded or very partial. Consideration should always be given to first improving the quality of the data biologically, for example by replicate PCR, DNA clean up or re-extraction of the sample before efforts to examine the profile statistically. Note that increasing the number of accepts will increase the run time.

Selecting **Cancel** in **Run Settings** during STRmix™ analysis set-up will return the user to the previous screen. Data entered on the previous screen will be retained. Selecting **Cancel** in STRmix™ analysis set-up will return the user to the Main Menu. All previously entered data will be cleared.

3.3.2 Add evidence profile data

Step	Action
1	Select the appropriate kit from the drop down list next to **Profiling Kit.** (Figure 3.6 Add Profile Data window)
2	Select **Kit Settings** in order to edit settings for this run (refer Figure 3.7). Refer to section 7 for an explanation of the settings. Note that Kit Settings have been separated into three tabs: General Settings (Figure 3.7 Kit settings/General settings), Stutter Settings (Figure 3.8 Kit settings/Stutter settings) and Loci Settings (Figure 3.9 Kit settings/Loci settings). Select **Apply** to save changes and return to the Add Profile Data window or **Cancel** to return without saving changes.
3	Evidence input files may be entered by 'drag and drop' into the Evidence profile data window or by navigation to the file path using the **Add evidence** ("**+**") button (see Figure 3.6). STRmix™ files and text files containing only one profile (ignoring any ladder) will be added automatically. If a text file contains more than one sample (or if you have selected to Allow Text Sample Renaming as a STRmix™ Default in the Administration area) the Add Evidence Profile Data window will open (see Figure 3.10). To add multiple profiles from the one file (e.g. replicate evidence input files), tick the selection box to the left of each sample you wish to add or use the Select All tick box at the bottom left of the window.
4	If adding profiles by navigation, in the Add Evidence Profile Data window choose **Select** and navigate to the input file. Select Add Profile Data to add the input file and **Confirm** or **Cancel** to return to the Add Profile Data window. File names may be edited if this feature is enabled within STRmix™ Defaults.
5	Repeat steps as above for any replicate evidence input files as appropriate.

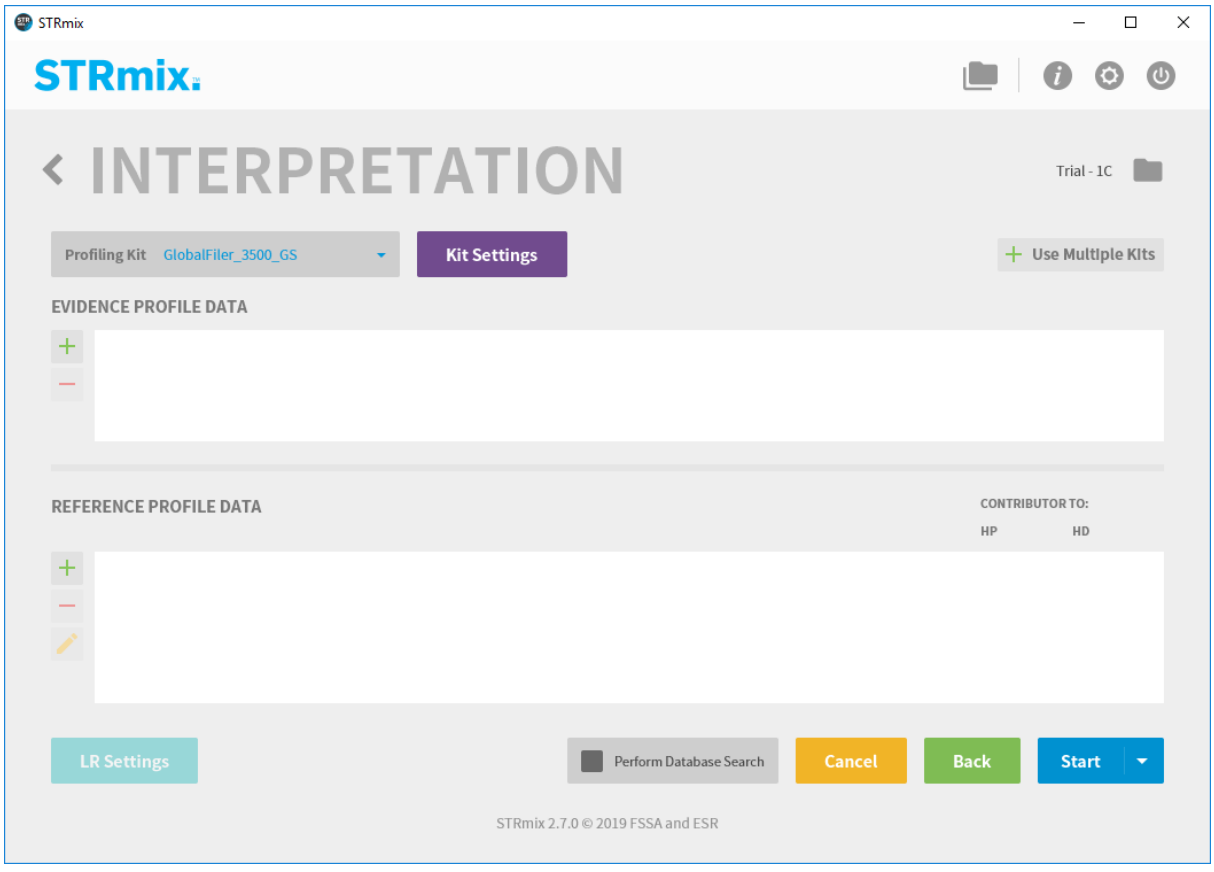


Figure 3.6 Add Profile Data window

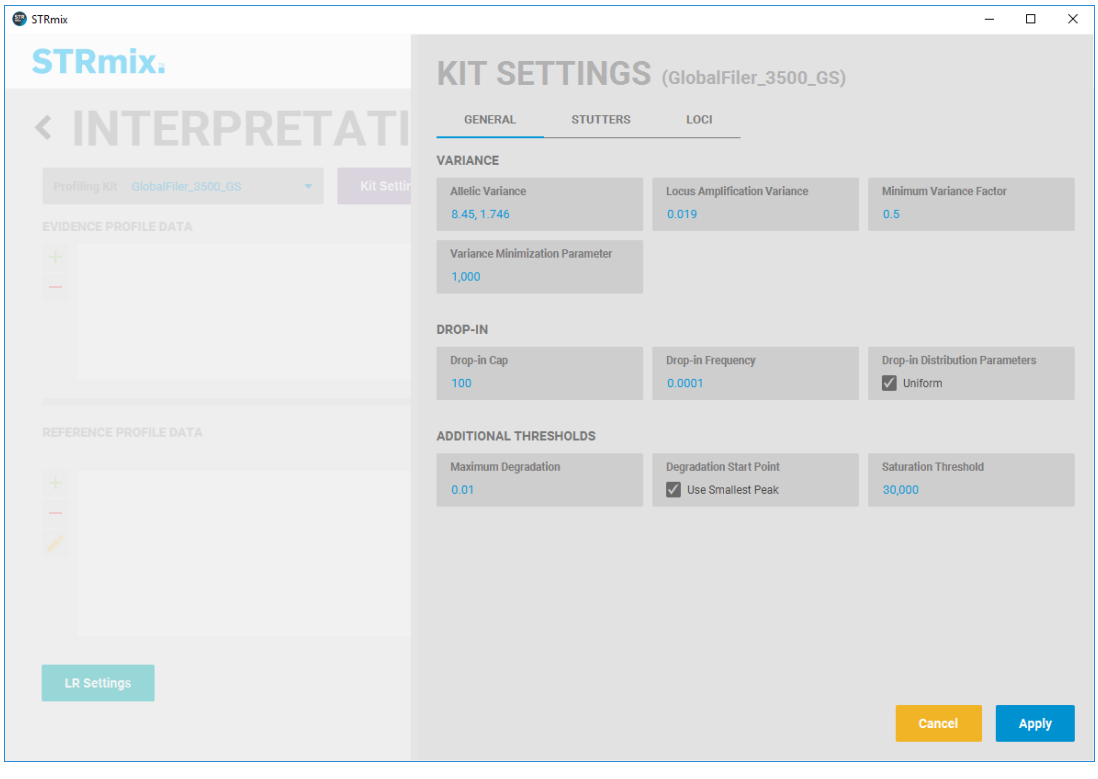


Figure 3.7 Kit settings/General settings

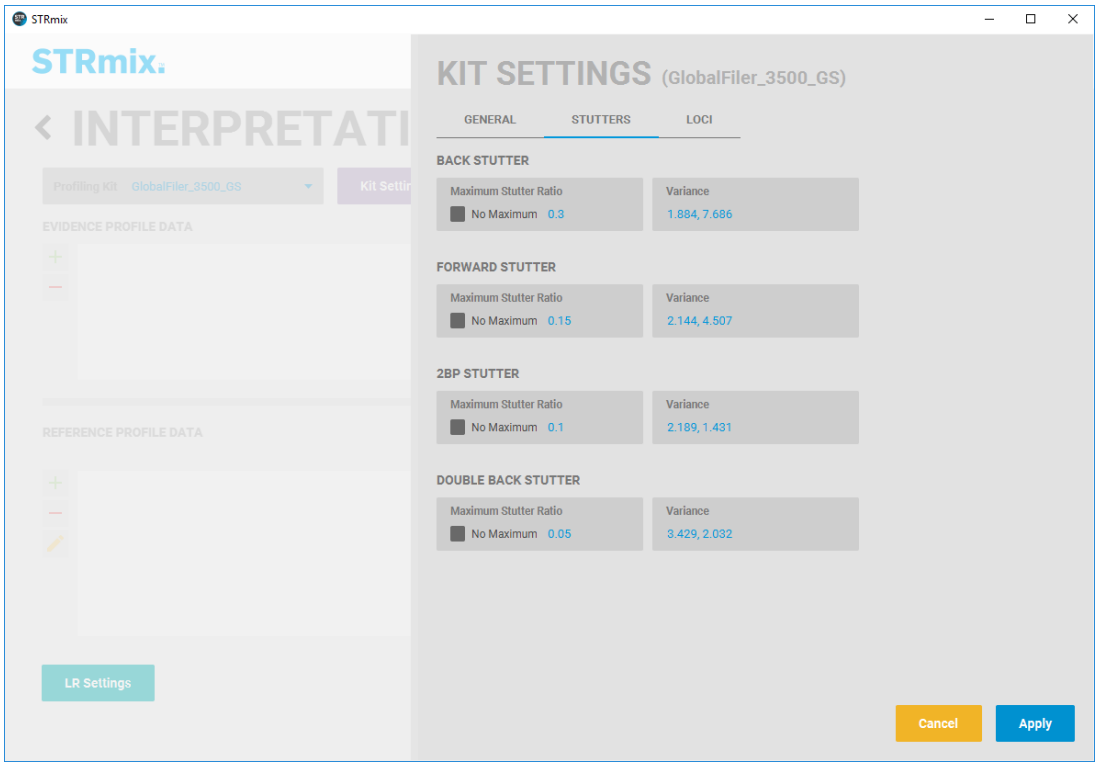


Figure 3.8 Kit settings/Stutter settings

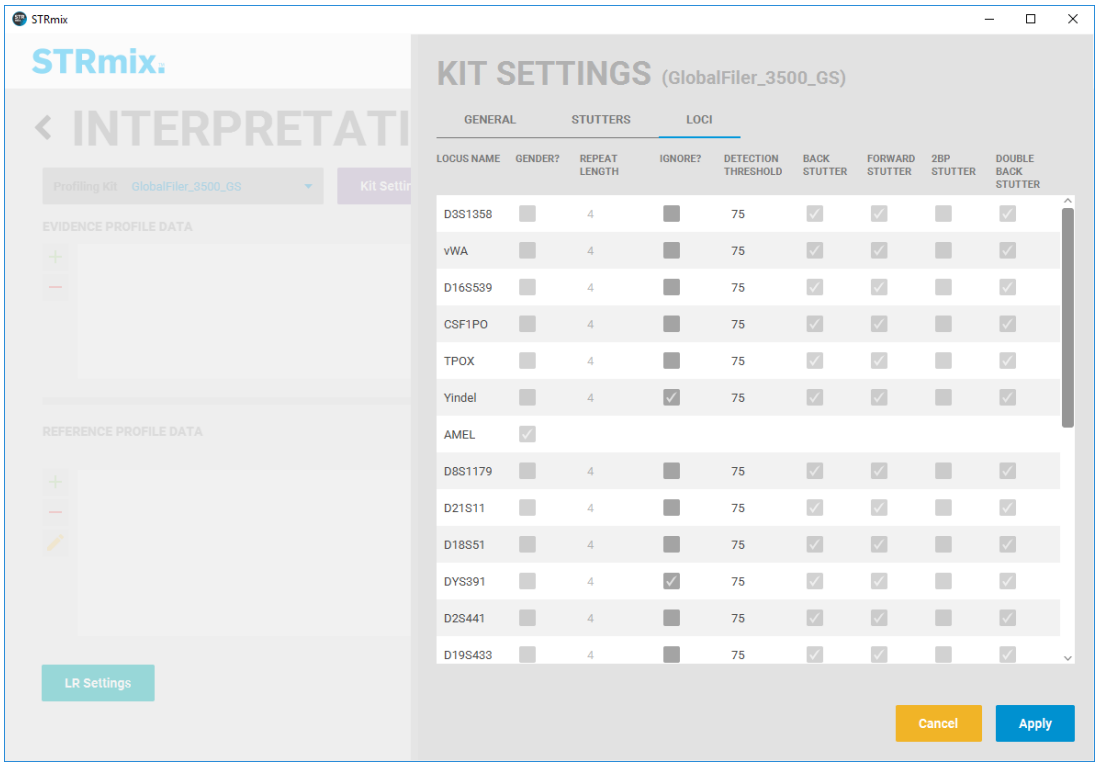


Figure 3.9 Kit settings/Loci settings

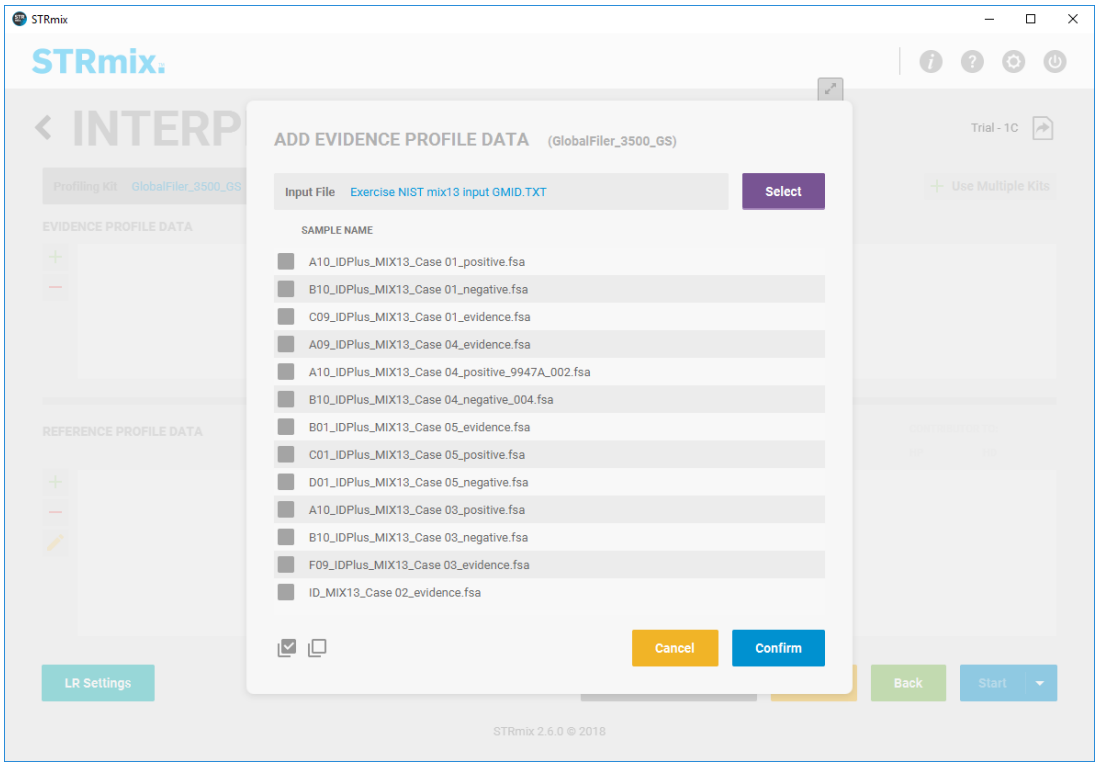


Figure 3.10 Add evidence profile data window

Notes

Ignore Loci: This function should be used rarely. An example of when this function might be used is when a reference sample profile is known to have a triallelic pattern.

The DNA Profiling kit used can be changed by selecting the appropriate value from the drop down box. STRmix™ is supplied with a number of pre-configured DNA kit exemplars. Additional kits may be created in the Administration > Profiling Kits window (refer section 7.3.1). Analysis of an Evidence epg/s can proceed without the need to enter any Reference epg input files.

To condition on an individual, highlight a selected reference file from the Add Profile Data window (Figure 3.6) and tick the **Hd** box (see 3.3.4 below).

To remove a profile, highlight a selected input file from the Add Profile Data window (Figure 3.6), and select **Remove Profile** (the "**-**" button) to delete the file.

3.3.3 Adding profiles from multiple kits

Step	Action
1	Select the **Use Multiple Kits** button from the right hand side of the Add Profile Data window (Figure 3.6) and choose a second kit from the drop down menu then **Select (**or **Cancel)**.
2	Highlight the kit you wish to change settings for and select **Kit Settings** to open the window in order to edit settings for this kit. Select **Save** to save changes and return to the Add Profile Data window or **Cancel** to return without saving changes.
3	Add the evidence files by drag and drop or navigation to the source folder as described above.

4	Repeat steps as above for any replicate evidence input files as appropriate.
5	Complete the steps above for as many kits you have profiles for.

3.3.3.1.1 Notes

Select Fix degradation to fix the degradation values between multiple kits. If enabled, degradation will be set to the first kit. If disabled, degradation will be optimised within the MCMC for both kits.

There is no limit to the number of kits that can be added.

If conditioning on an individual within a multi kit interpretation, the reference must contain all loci across the different kits. If not, ignore the missing loci from the interpretation.

The multi kit logic assumes the input files were generated from the same template (DNA extract).

In V2.8, multi-kit analysis is enabled for Interpretations using a range of contributors (varNOC).

3.3.4 Adding reference profiles

Step	Action
1	In the Reference Profile Data box add text or STRmix™ input reference input files as above by drag and drop or navigate via buttons.
2	Reference input files will automatically be assigned as a known contributor under H_p. Should you also require a reference sample to be an assumed contributor under H_d then select the reference of interest using the cursor so it is highlighted then tick the **Hd** box.
3	To edit a reference, highlight the sample of interest and select the **pencil icon**. Reference names may be edited and a *stitched* reference may be created by adding multiple files from the same individual. A concordance check is undertaken for any overlapping loci. When adding multiple files for the same reference the reference name will automatically be appended with the suffix “CREF” alternatively you may choose to edit the name yourself. Select **Save** to continue or **Cancel** to return to the Add profile data window.
4	Select **Start** to proceed (refer Figure 3.11), **Cancel** to return to the Start menu or **Back** to return to the Case Details window. Selecting **Back** will not remove profiles previously uploaded.

Notes

A mixture can be deconvoluted with an assumed contributor without calculating an *LR* by setting up STRmix™ as in Figure 3.12. This may be useful for a case with no reference samples from POIs where an individual contributor may be assumed (e.g. the complainant present in their own intimate sample).

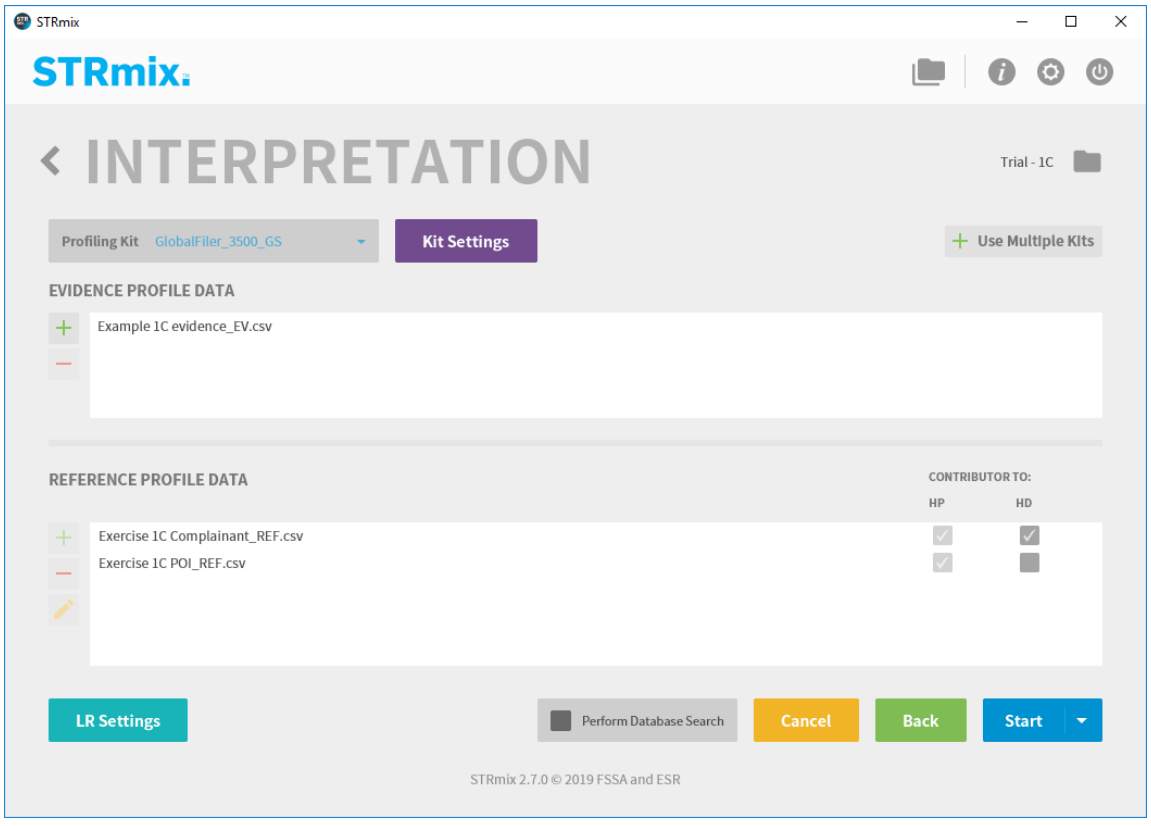


Figure 3.11 STRmix™ setup for an example with one assumed contributor and one POI

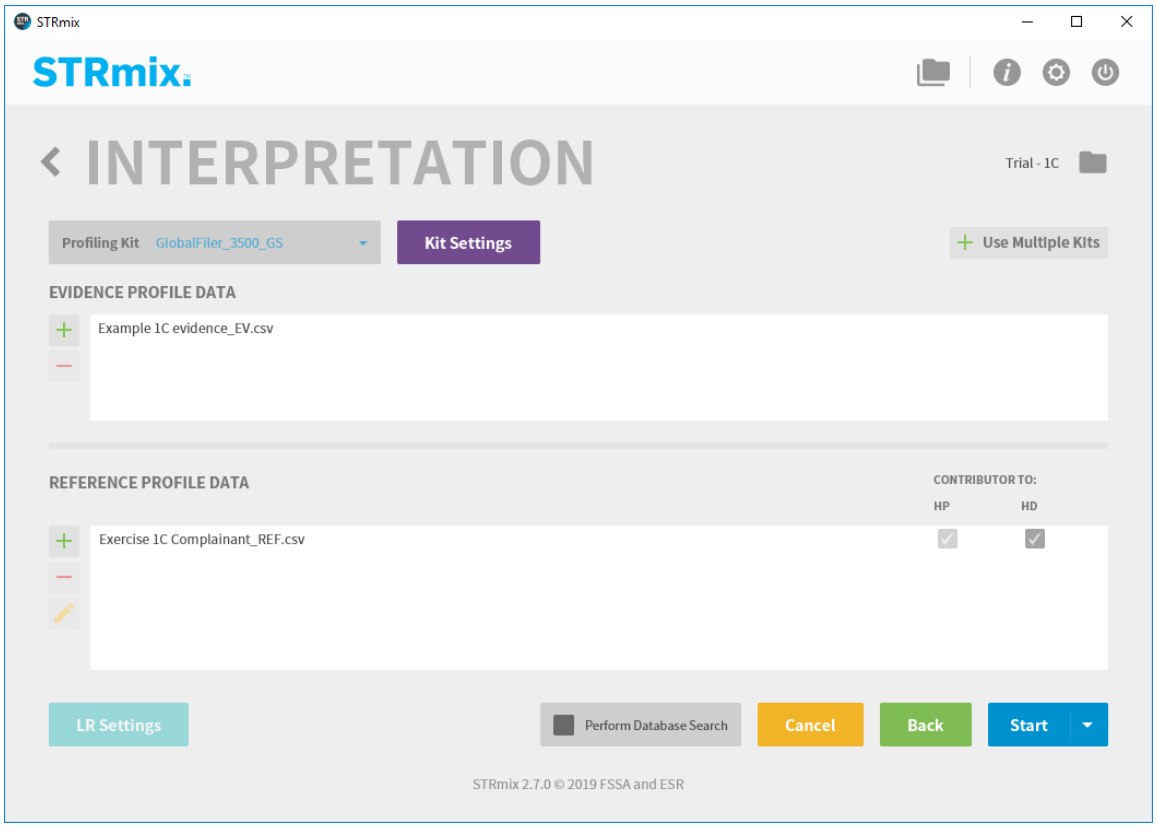


Figure 3.12 STRmix™ setup for an example with one assumed contributor and no POI

3.3.5 Selecting population settings

Step	Action
1	Select LR Settings. Select **"+"** to add a population. Select the population of interest from the list. Hold down the Control key or Shift key to select multiple populations. (Refer to section 7.4)
2	Check Proportion and F_{ST} values. F_{ST} values can be edited for the run by selecting the F_{ST} value to alter, entering the new value and selecting **OK** to change (or **Cancel** to cancel).
3	Select the disk icon to save the population settings as the default values for this kit. If enabled, the Administrator password is required for this change.
4	Check Sampling Variation parameters, Sub-source LR, and Contributor Range LR method (see section 3.3.6) as required. Contributor Range priors will only be available if a range of contributors has been selected on the Interpretation set-up screen. Select **Apply** to save the changes or **Cancel**.
5	Select **Start** to start the calculation, **Cancel** to return to the Start menu or **Back** to return to the Interpretation set-up window. Check the **Perform Database Search** option to start the calculation followed by an automatic search against a preselected database (refer to section 4.2.3 for details). Note that default database search parameters must be saved first (refer Section 7.1).

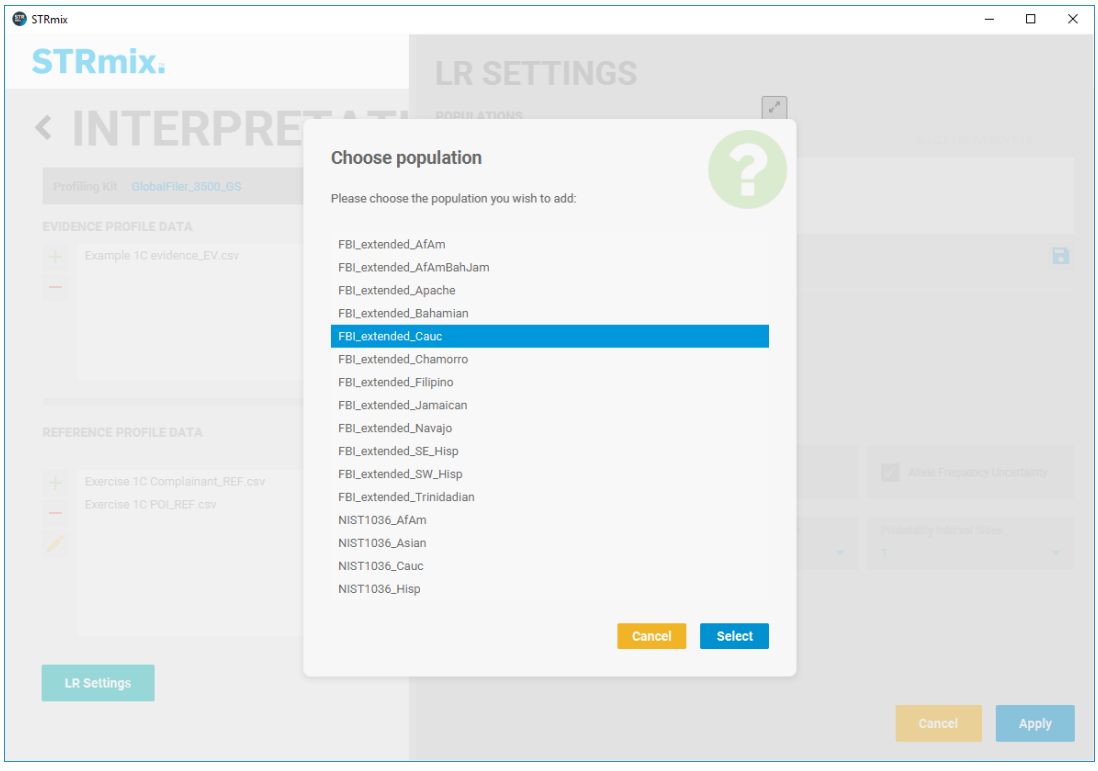


Figure 3.13 Choose population

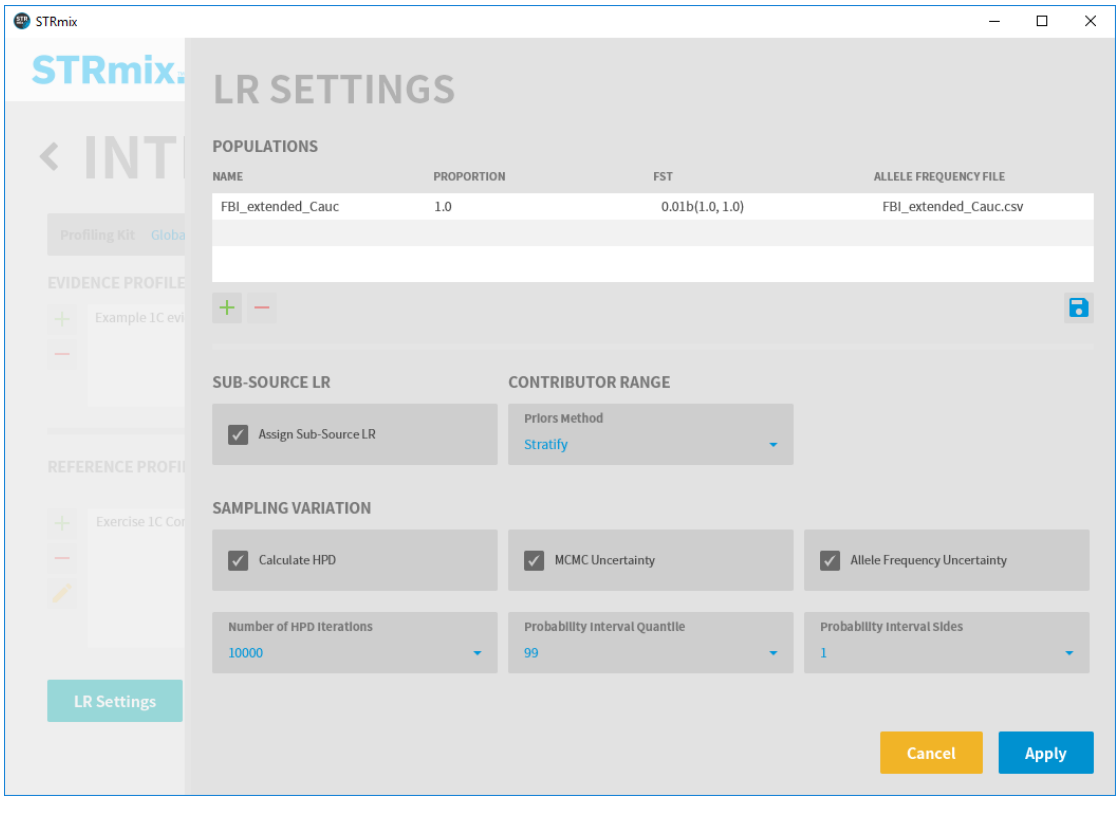


Figure 3.14 LR settings

3.3.6 Selecting Contributor Range LR Method

When carrying out an *LR* interpretation with a range of contributors (varNOC), there are three methods available within the *LR* settings for assigning an *LR* – Stratified *LR* (Stratify), Maximum Likelihood Estimate *LR* (MLE) and User defined *N* under H_p and H_d (User selected).

3.3.6.1 Stratified *LR*

Selecting **Stratify** from the **Priors Method** dropdown within LR Settings will assign a weighted average *LR* across the numbers of contributors.

3.3.6.2 Maximum Likelihood Estimate *LR* (MLE)

Selecting **MLE** from the **Priors Method** dropdown within LR Settings will assign an *LR* using the most probable number of contributors under H_p and H_d.

3.3.6.3 User defined N under H_p and H_d

Selecting **User Selected** from the **Priors Method** dropdown within LR Settings allows the user to choose the number of contributors under H_p and H_d, within the assumed number of contributors range under which the profile was interpreted (see Figure 3.15).

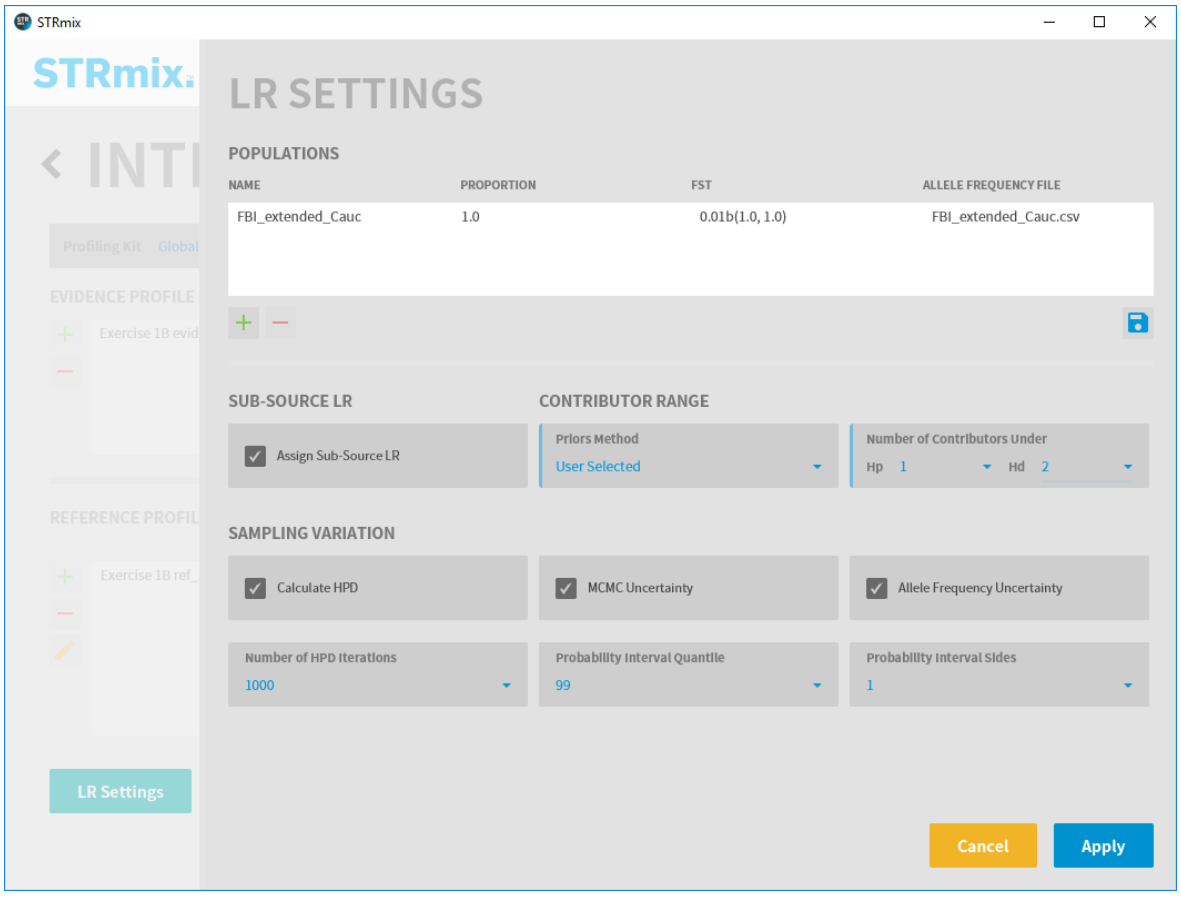


Figure 3.15 Setting User defined N under H_p and H_d

Notes

Population settings are only required when reference samples have been added (not conditioning). If no reference samples are selected, then no population is required. Select **Start** to start the profile interpretation.

To use the **Database Search** option, a default database must be saved in Administration > STRmix™ Defaults > Database Search (refer section 7).

STRmix™ is supplied with a number of pre-configured exemplar populations. Additional populations may be created in the Administration > Populations window (refer section 7.4.1). If multiple populations are selected then a stratified *LR* value may be provided as well as individual population values. Population proportions will automatically be normalised to sum to one.

To remove added populations, highlight the population and select "**-**".

To change the F_{ST}, click the F_{ST} value to open the Edit FST field. Type in the new F_{ST} parameters and select **OK**. Note that the revised F_{ST} is only changed for the current run (unless the save icon is selected: ).

By unchecking the **Calculate HPD** box, only a point estimate *LR* is returned. Alternatively, the **MCMC Uncertainty** or **Allele Frequency Uncertainty** can be disabled from the HPD calculation.

By unchecking the **Assign Sub-Source *LR*** box, the sub-sub-source *LR* is calculated and no relative *LR*s are calculated [1].

The default values for the number of HPD iterations, significance level and sides to the test are configured within the Sampling Variation fields. These parameters can be changed in Step 4 for the current run.

To save a zipped folder of the interpretation setup, select Export from the dropdown list from the Start button. Select the file location to save the folder. You will be returned to the Main Menu and the interpretation may be started at a later time.

3.3.7 Version concordance

Within STRmix™ V2.8, reference files are not associated with a kit. If you select an indexed STRmix™ input file (where loci are defined by locus numbers and not names) you will be required to select the relevant kit on entry. A dialog box will ask for the appropriate kit file to be selected (see Figure 3.16).

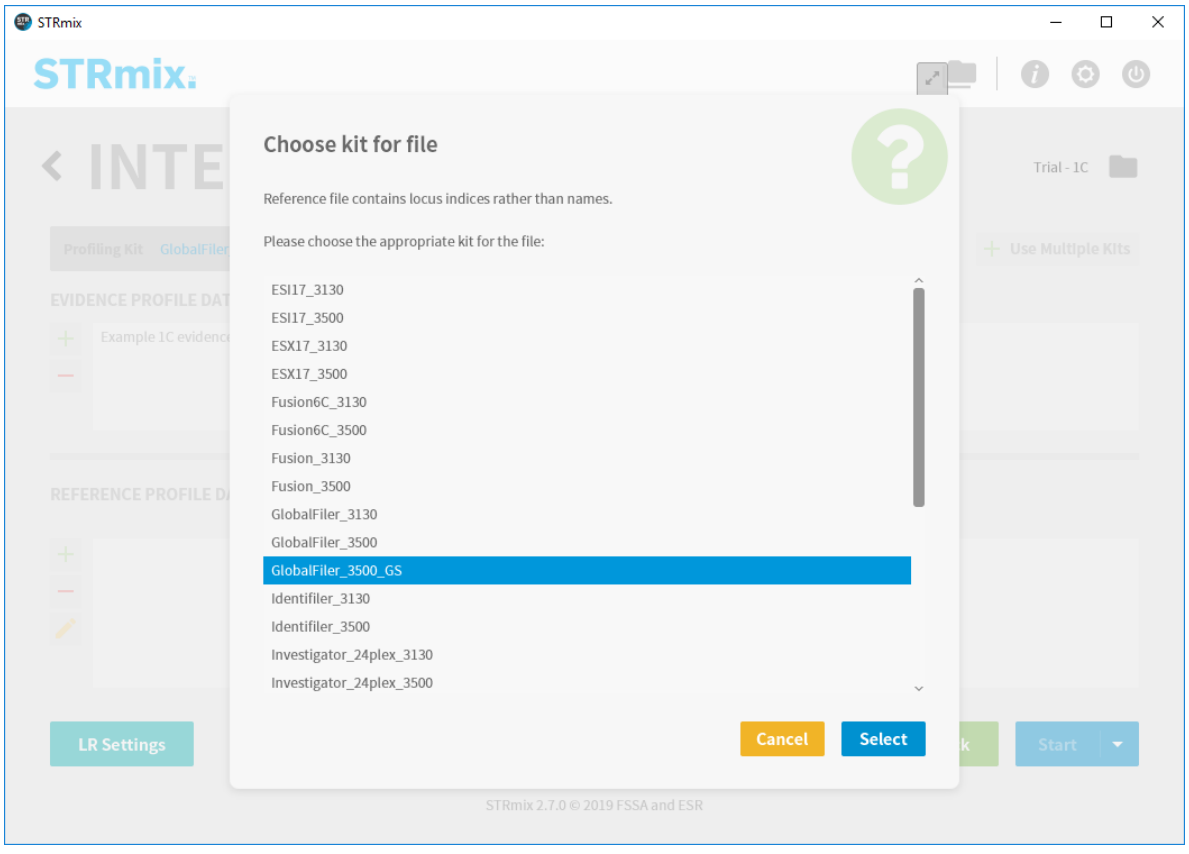


Figure 3.16 Kit file selection window when indexed STRmix™ input file is added

3.3.8 User informed mixture proportion priors

The default settings in STRmix™ use uninformed priors for DNA amount and hence mixture proportions (Mx). STRmix™ has the ability for the user to provide informed priors through the use of the **Use Mx priors** option prior to interpretation. There are some instances when a user may wish to provide informed priors for Mx. One instance would be when information is present in an epg that is below the detection threshold, suggesting a low level contributor. In these circumstances, as the information has not been detected, STRmix™ will not be able to make use of that information. The user can overcome this by choosing an informed Mx for one contributor at trace levels.

Another instance where informed Mx priors may be used is if the user has scenario specific information that multiple relatives have contributed DNA to a sample. In this situation profiles will have significant allele overlap and there may be a number of Mx solutions that fit the profile. In this instance the user may wish to use the case context to provide informed Mx priors.

The informed prior is set before the interpretation, after reviewing the epg. Mx priors are set from the Run Settings screen (refer

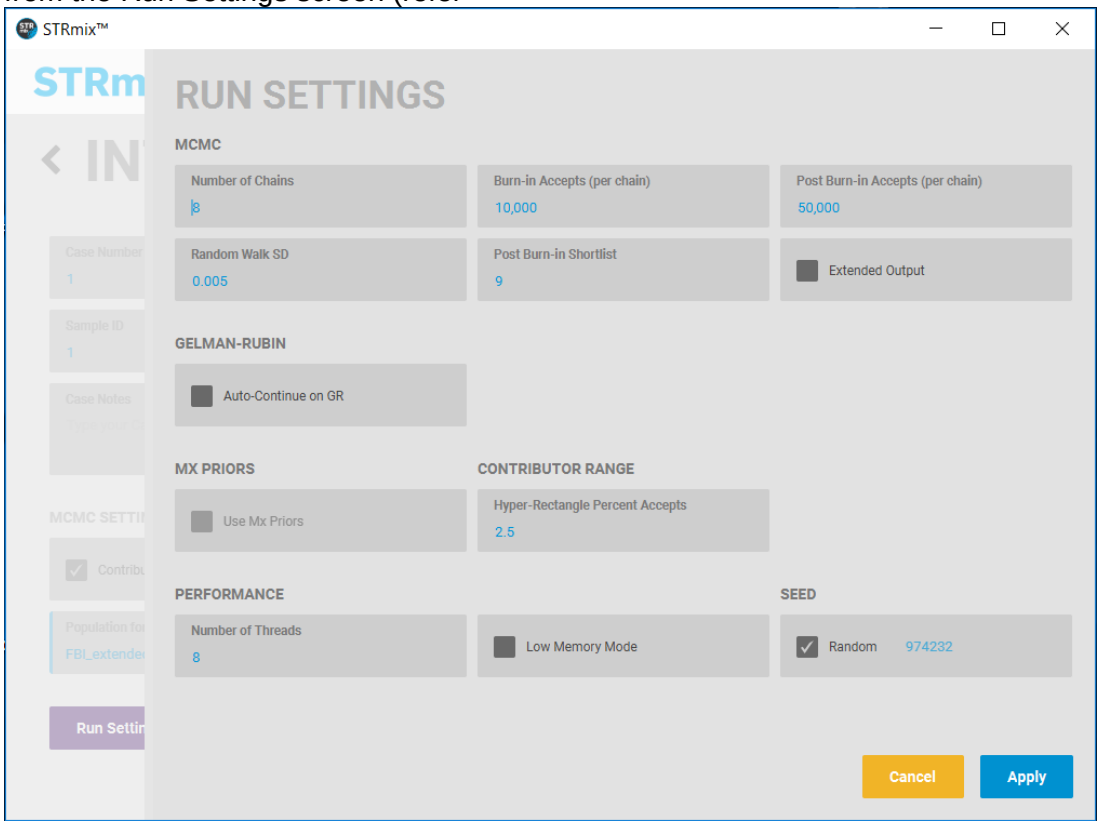


Figure 3.5). In the Interpretation Case Details screen, select **Run Settings.** In the Run Settings screen, select **Use Mx priors**. The window in Figure 3.17 will open. Select the contributor in the drop down box. Slide the mean bar to select the mean of the Mx and var bar for the variance for each contributor. Priors for any assumed contributor/s must be set first, otherwise they should be set in order of decreasing mixture proportion.

As an example, in Figure 3.17 the prior mean for contributor 1 is 0.66 and for contributor 2 is 0.27. They have the same variance. In Figure 3.18 the prior mean for contributor 1 is 0.66 and 0.27 for contributor 2. The variance for contributor 1 is a lot higher than for contributor 2 indicating the uncertainty in this mean. The contributors should be added in order of highest to lowest proportion. If you are assuming a known contributor this should be in the contributor 1 position. Note the proportions do not need to sum to 1.0.

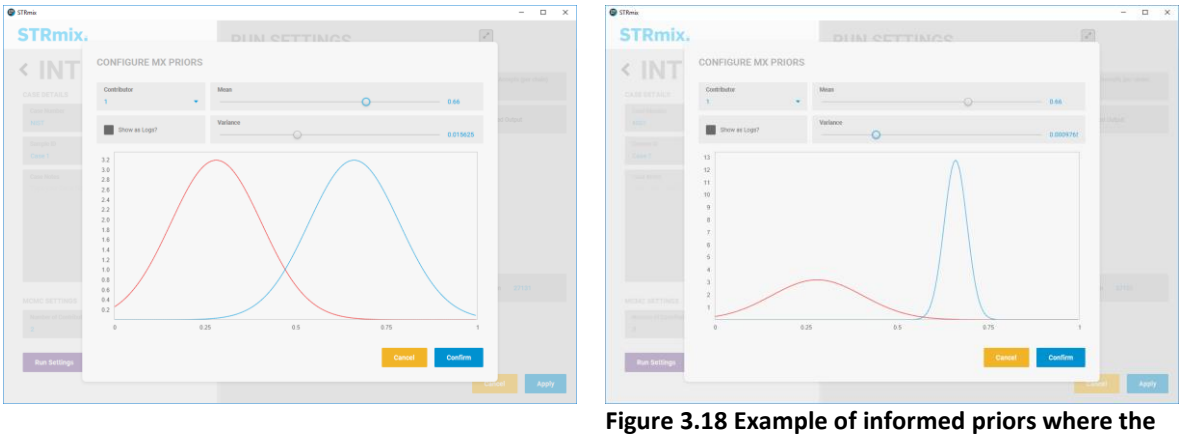


Figure 3.18 Example of informed priors where the variance for contributor 2 is higher than contributor 1

Figure 3.17 Example of informed priors where contributor 1 and 2 have the same variance

Selecting and deselecting the **Show as Logs?** checkbox will cycle the *y*-axis scale on the plot between densities and log densities. This is visual only and will not affect the calculation, as both graph types display the same priors, just on different scales.

3.3.9 Stutter and Analytical Threshold pre-checks

When **Start** is selected STRmix™ will perform quality checks on the input data to check for missing stutter peaks and peaks below the analytical threshold set in the kit.

3.3.9.1 Analytical Threshold

If peaks in the evidence input file are found to be below the analytical threshold of the kit selected, the user will be prompted to stop the interpretation, or continue the interpretation but ignore the peaks that are below the analytical threshold, or continue the interpretation as it is (Figure 3.19).

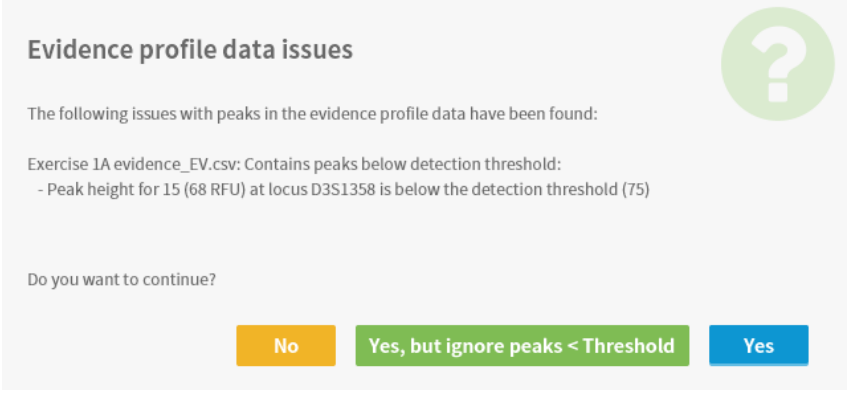


Figure 3.19 Warning message for peaks below the analytical threshold.

If **No** is selected, the user is returned to the Interpretation screen to investigate the evidence input file or Cancel the Interpretation.

If **Yes, but ignore peaks < Threshold** is selected, the Interpretation will progress (see 3.3.10 Running) and a message will be written out to the log and the Evidence Peak Issues section of the pdf report (see Figure 3.20).

If **Yes** is selected, the Interpretation will progress with all input peaks included and a message will be written out to the log and pdf report as above (see Figure 3.21).

EVIDENCE PEAK ISSUES

EXERCISE 1A EVIDENCE_EV.CSV

LOCUS	PEAK	ISSUE	DECISION
Peaks below Detection Threshold			
D3S1358	15	Peak height for 15 (68 RFU) at locus D3S1358 is below the detection threshold (75)	Peak Ignored

Figure 3.20 Evidence peak issues from pdf report for peaks below analytical threshold where "Yes, but ignore peaks < Threshold" was selected

EVIDENCE PEAK ISSUES

EXERCISE 1A EVIDENCE_EV.CSV

LOCUS	PEAK	ISSUE	DECISION
Peaks below Detection Threshold			
D3S1358	15	Peak height for 15 (68 RFU) at locus D3S1358 is below the detection threshold (75)	Peak Retained

Figure 3.21 Evidence peak issues from pdf report for peaks below analytical threshold where "Yes" was selected

3.3.9.2 Missing stutter peaks

Missing stutter peaks can lead to large negative log(likelihood) values, large stutter variances, the non-convergence of parameters between different chains, and in some cases false exclusions of true contributors. A number of these missing peaks may be explained by non-resolution of 1 bp variants. Within STRmix™ V2.8 a pre-check for missing stutter peaks is undertaken prior to the interpretation starting.

For each peak within the profile, the input file is checked for the presence of all stutter types modelled within the interpretation. For each missing stutter peak the dropout probability is calculated. For back stutter peaks this is:

$$\log\left(\frac{z}{E_{a-1}}\right) \sim N\left(0, \frac{\sigma^2}{O_a}\right)$$

Where z is the analytical threshold, E_{a-1} the expected height of the stutter peak, σ^2 the mode of the back stutter variance prior, and O_a the observed height of the putative allele. If the dropout probability is greater than a pre-determined threshold (default -8) the interpretation progresses. If the dropout penalty is smaller than the threshold the interpretation is paused and a user is prompted to continue or halt the deconvolution (see Figure 3.22).

If the interpretation is resumed with missing stutter identified, a comment is written to the report and log file (see Figure 3.23). In Batch Mode the interpretation progresses automatically and any missing stutter is written to the report and log file.

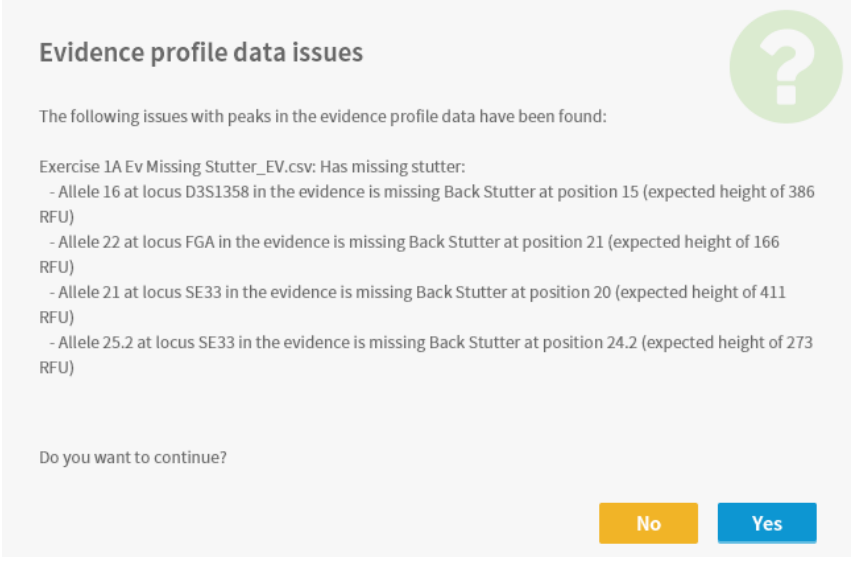


Figure 3.22 Warning message for missing stutter

EVIDENCE PEAK ISSUES

EXERCISE 1A EV MISSING STUTTER_EV.CSV

LOCUS	PEAK	ISSUE	DECISION
Missing Stutter Peaks			
D3S1358	15	Allele 16 is missing Back Stutter at position 15 (expected height of 386 RFU)	-
FGA	21	Allele 22 is missing Back Stutter at position 21 (expected height of 166 RFU)	-
SE33	20	Allele 21 is missing Back Stutter at position 20 (expected height of 411 RFU)	-
	24.2	Allele 25.2 is missing Back Stutter at position 24.2 (expected height of 273 RFU)	-

Figure 3.23 Evidence peak issues from pdf report for missing stutter peaks.

3.3.10 Running

Following pre-checks, STRmix™ will start the calculation. A Calculation Progress window displaying the date and time the STRmix™ calculation began will open. The calculation progress will be indicated in this window with run progress visualised by a progress bar (see Figure 3.24).

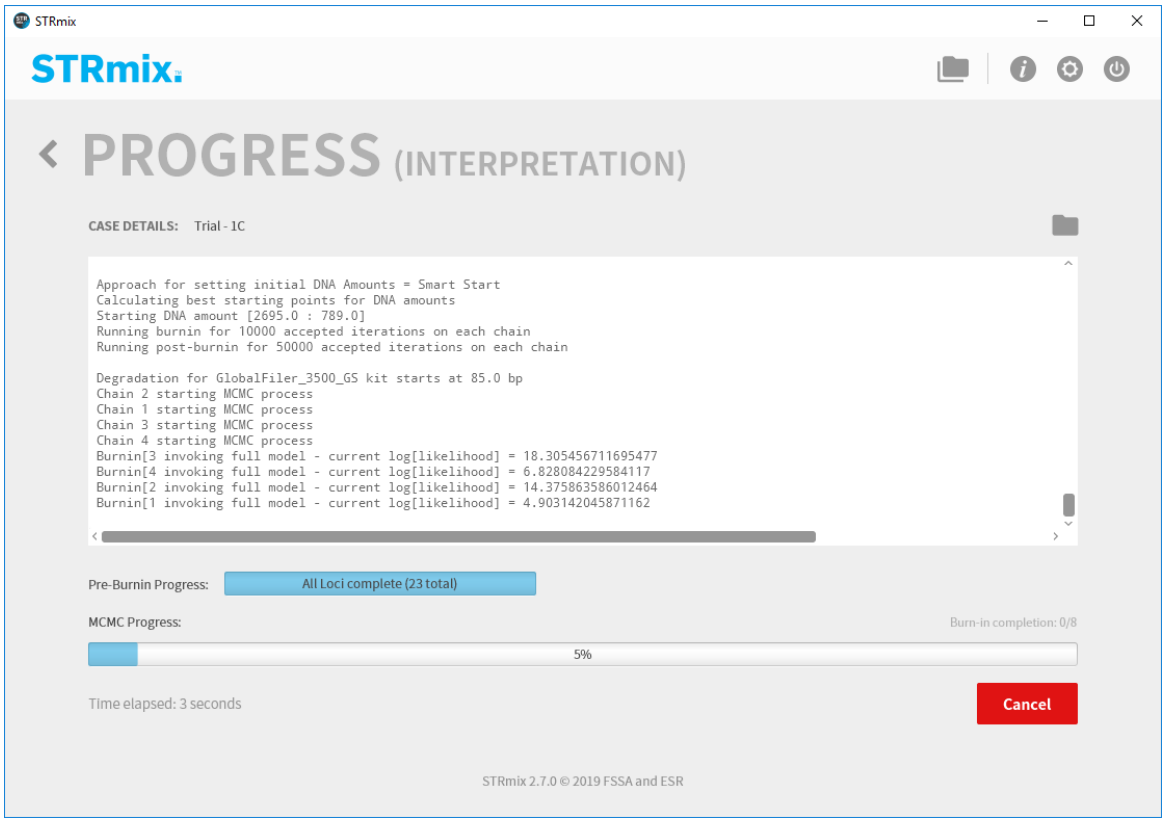


Figure 3.24 Calculation progress window

When the calculation is finished a summary of the findings of the analysis will open automatically (see

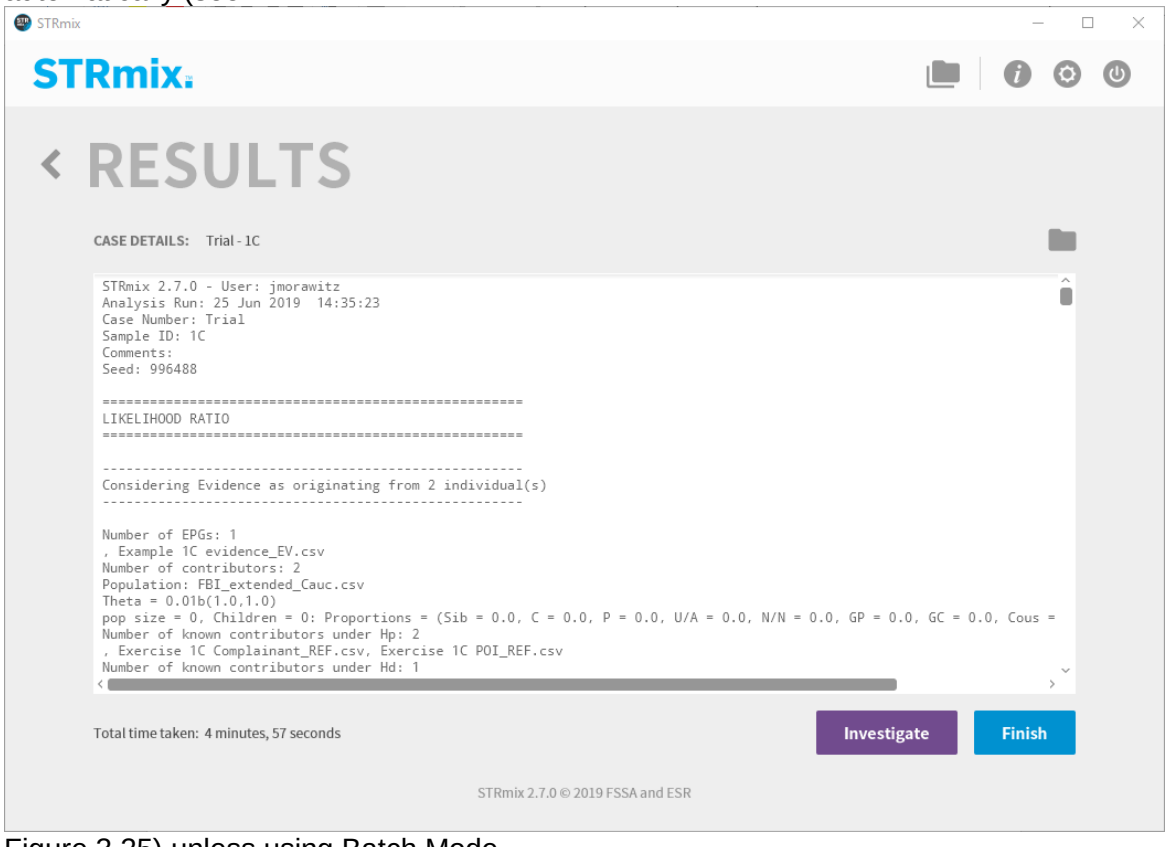


Figure 3.25) unless using Batch Mode.

3.3.11 Results

The results of the interpretation (and investigation, described below) are saved automatically in a new folder. The Results folder for the completed run may be opened by selecting the Open Interpretation Results Directory icon:  in the upper right third of the Summary window (see Figure 3.25). Select **Finish** to return to the main window or **Investigate** to open the Investigation window and carry out further investigation of the results.

Within the specified Results directory a number of folders and files will have been created for the Interpretation. These are:

1. AlleleFreq folder containing each allele frequency used within the analysis (where appropriate)
2. Audit folder containing SHA512 hash information for the files used and generated in the analysis (see Section 9)
3. Extended output folder containing size regression information and any extended outputs where appropriate
4. Inputs folder containing all the input files saved in STRmix™ format (appended _EV or _REF or _CREF as appropriate)
5. Kits folder containing all kit files used within the analysis
6. Log folder containing the Interpretation output printed within the calculation progress window (see Figure 3.24)
7. Populations folder containing each population file used within the analysis (where appropriate)
8. Reports folder containing the PDF report and reporting elements
9. Stutters folder containing all stutter and size regression files used within the analysis
10. A codisResults.xml where appropriate
11. A config.xml and config_input.xml file, variously containing all run information
12. The Component Interpretation containing all accepted genotypes and weights per contributor
13. The Genotype probability distribution produced by the analysis containing all accepted genotypes and weights
14. A Results.txt file containing result information including the ten genotype combinations at each locus with the highest weights.
15. A results.xml file containing result information

Refer to Section 6 for a full explanation of the reports.

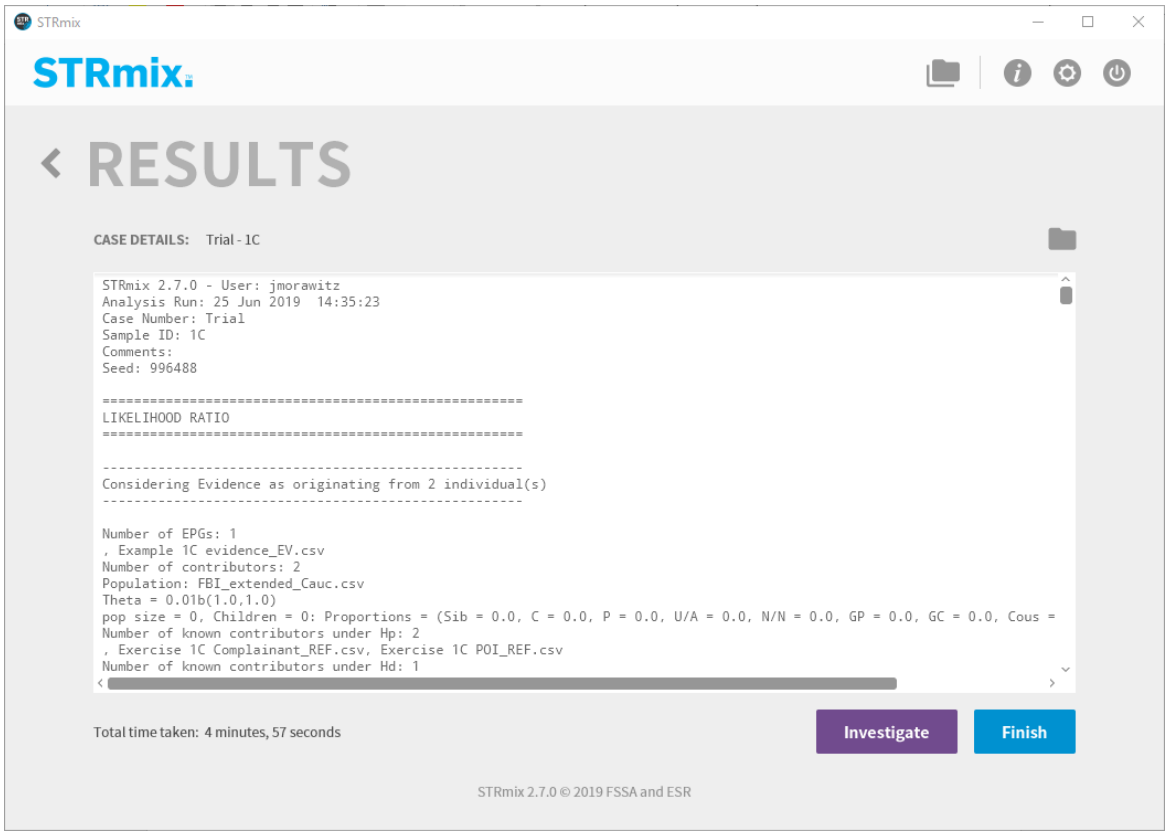


Figure 3.25 Summary of results at completion of STRmix™ run

A copy of the results may be automatically saved to a second location, for example a central location if STRmix™ is installed to individual analyst's computers. The settings are configured by adding the following text to the Default.ini file within the STRmix™ directory:

CopyResults

Y

CopyResultsDir

D:\Some\Path\To\Copy\To

Copying can be disabled while preserving the path to the copy directory by CopyResults to N. If CopyResults is enabled (Y) but no CopyResultsDir is set, an INTERPRETATION COPY FAILED.txt file will be written to the original (local) results directory. If for some reason the copy failed (e.g. due to a network error) the same INTERPRETATION COPY FAILED.txt file will be written to the local results directory, with details of the copy error.

4. INVESTIGATION

Within the Investigation tab from the Main Menu (refer Figure 3.2) a number of features are available to explore a STRmix™ deconvolution. From the Investigation window (Figure 4.1), **Browse** to the config.xml from the interpretation folder or drag and drop the folder directly into the Previous Interpretation field. If you select Investigate at the end of a run (refer Figure 3.25 above) the Investigation window will open with the Previous Interpretation field already populated. Select the Investigation feature to use from the buttons available. Each of these is discussed in turn in this section.

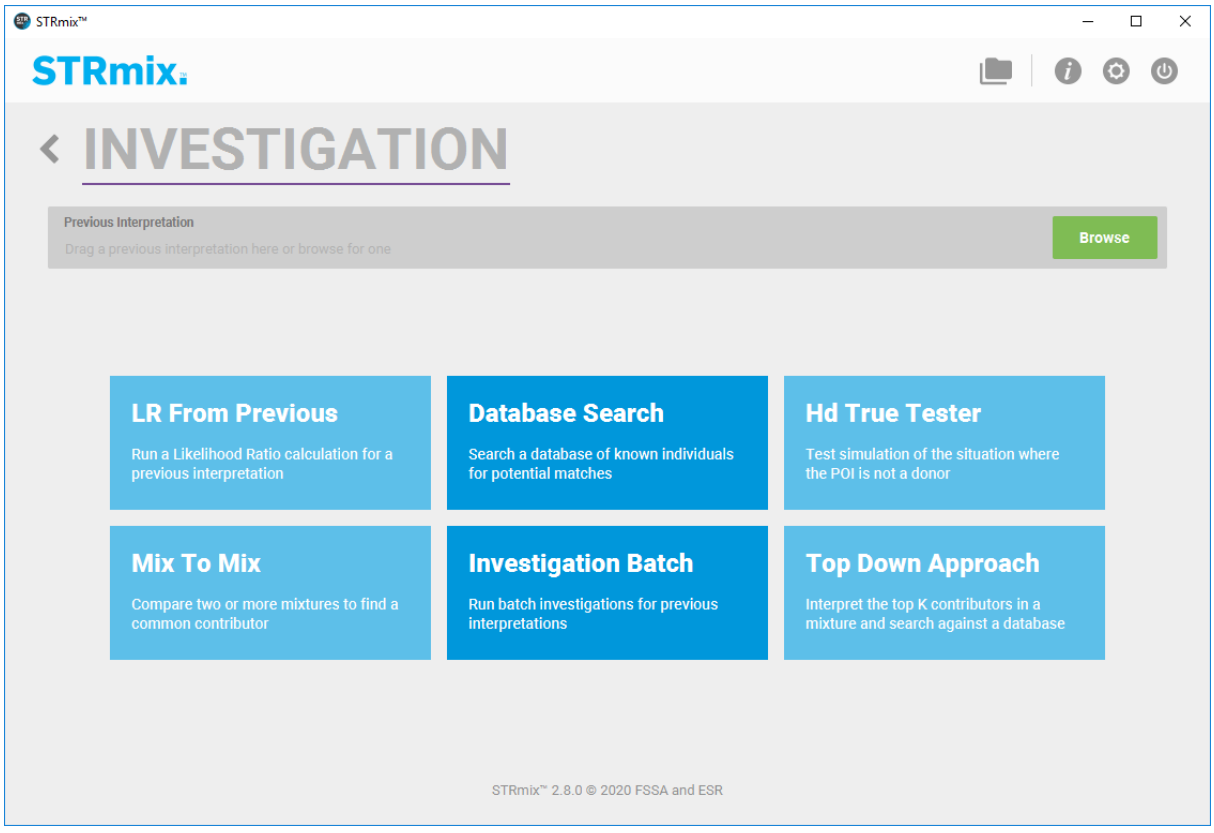


Figure 4.1 STRmix™ Investigation window

4.1 Calculating an LR from a Previous Analysis

If a mixture has previously been deconvoluted then the genotype probability distributions can be used to calculate an *LR* by comparing to additional references as they are obtained without the need to reanalyse the mixture. This function can be especially useful when reference samples from POIs become available several months or years after the original STRmix™ deconvolution was undertaken.

Select **LR from Previous** from the Investigation window (refer Figure 4.1). If you have not previously added a deconvolution within the Investigation window, the Choose Previous Interpretation for *LR* window will open. Drag and drop a results folder or choose **Browse** and navigate to a results folder to find the config.xml (or Settings.ini file if using a deconvolution from a version of STRmix™ prior to v2.5.11) from a previous interpretation (see Figure 4.2).

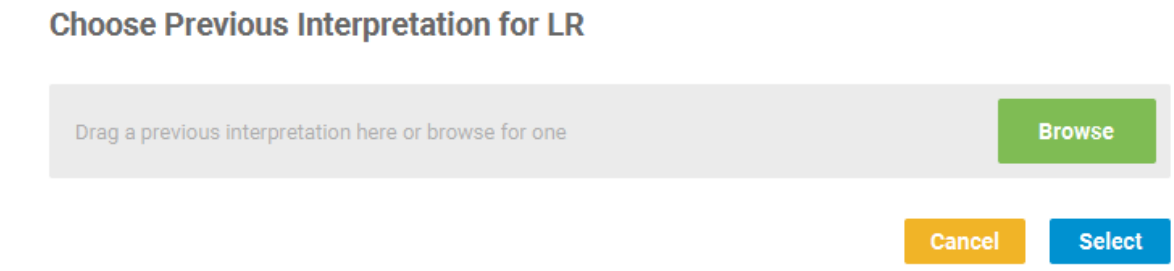


Figure 4.2 Choose previous interpretation for LR window for selecting a previous interpretation config.xml file (or Settings.ini)

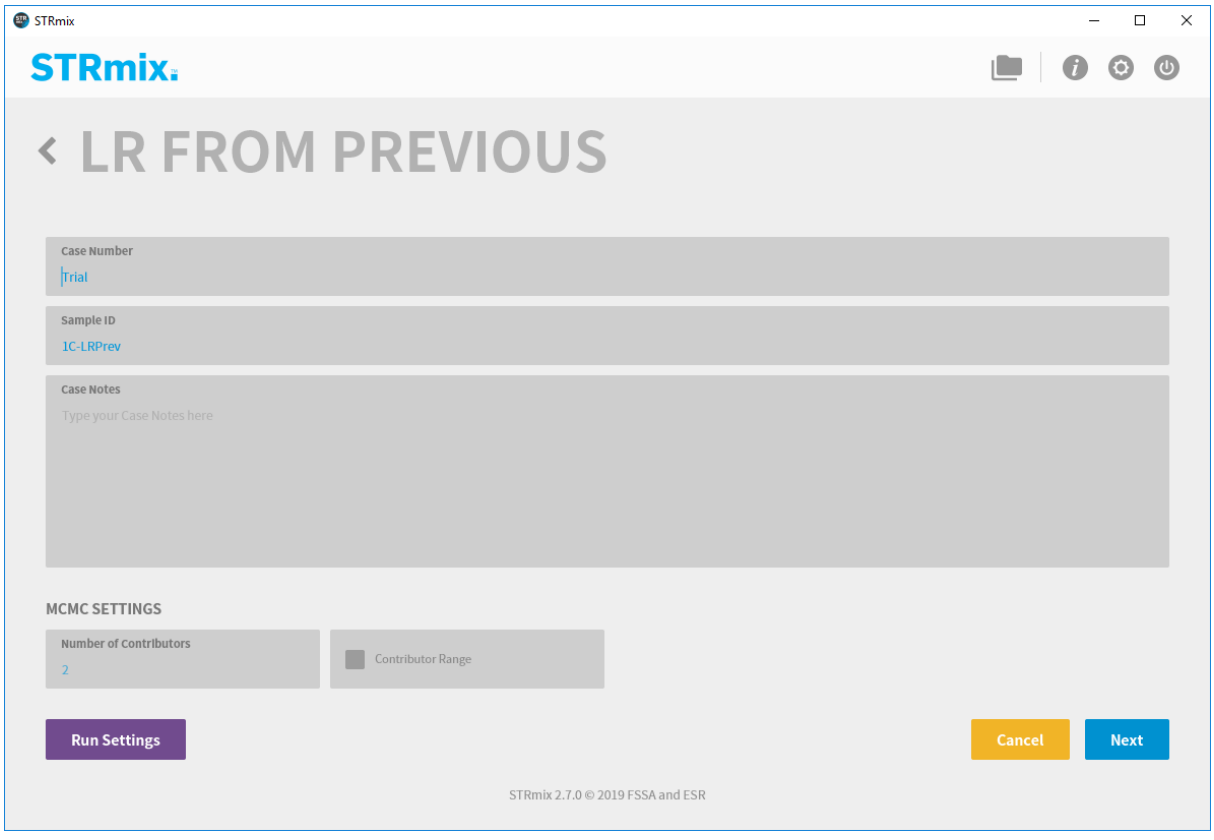


Figure 4.3 *LR* From Previous Window

The Case Number, Sample ID & Case Notes can be updated as required. Note that LRPrev is automatically appended to the Sample ID.

The only **Run Settings** enabled are the Extended Output, Low Memory Mode, and the Number of Threads and the Seed. The seed may be set by unchecking the **Random** box and entering a number. Setting the seed here for an *LR* from Previous will only affect the HPD *LR* results. Select **Apply** or **Cancel** to accept or cancel any run setting changes.

Select **Next** to proceed. Add the reference profile/s. The only **Kit Setting** that can be edited is ignore loci.

Select **LR Settings** to add or remove populations and change the *LR* settings (for example F_{ST}, HPD and Sub-Source *LR* on/off settings). Select **Apply** or **Cancel** to accept or cancel any *LR* setting changes.

Select **Start** to start the *LR* calculation.

Notes

Under ***LR* From Previous**, anything specific to the deconvolution itself cannot be changed, specifically; evidence epgs cannot be edited, deleted, or added and a reference epg cannot be an assumed contributor under both H_p and H_d. These scenarios can be performed by starting a new analysis from **Interpretation**.

The results summary output will not include the Genotype Probability Distributions from the original mixture deconvolution. If required, these can be obtained from the original run results folder.

4.2 Database Search

STRmix™ enables two potential strategies for database searching. These are

1. Using STRmix™ resolutions, and
2. Searching unresolvable mixtures.

4.2.1 Using STRmix™ resolutions

A possible strategy is to load to a database any genotype that STRmix™ assigns a weight greater than *x* at a locus. ESR and FSSA use *x* = 0.99. There is a risk/reward decision to be made. The risk of the uploaded profile containing at least one error across *L* loci using a value for *x* is $(1-x)^L$ shown for some values in Table 4-1 below[1]. Note that these values represent a worst case scenario where every genotype being interpreted has exactly a weight of *x*.

Table 4-1 Risk of at least one error given different match stringencies and number of loci

The risk of at least one error		*1-x*				
		0.01	0.02	0.03	0.04	0.05
L	9	9%	17%	24%	31%	37%
	10	10%	18%	26%	34%	40%
	11	10%	20%	28%	36%	43%
	12	11%	22%	31%	39%	46%
	13	12%	23%	33%	41%	49%
	14	13%	25%	35%	44%	51%
	15	14%	26%	37%	46%	54%
	16	15%	28%	39%	48%	56%
	17	16%	29%	40%	50%	58%

[1] We acknowledge an unlikely assumption of independence

	18	17%	30%	42%	52%	60%
	19	17%	32%	44%	54%	62%
	20	18%	33%	46%	56%	64%
	21	19%	35%	47%	58%	66%
	22	20%	36%	49%	59%	68%
	23	21%	37%	50%	61%	69%
	24	21%	38%	52%	62%	71%
	25	22%	40%	53%	64%	72%
	26	23%	41%	55%	65%	74%

This approach may also be applied to an allele. If two or more different genotypes both contain the same allele and collectively add to a weight greater than *x* then that allele may be loaded.

4.2.2 Searching unresolvable mixtures

STRmix™ offers the option of searching an unresolvable mixture against a database in Excel format. By unresolvable we mean where there is no apparent major or minor component, for example a 1:1 mixture. Each of the individuals on the database are considered as a potential contributor in turn to the mixture under the following two hypotheses:

H_p: Database individual and *N* - 1 unknown contributors

H_d: *N* unknown contributors

where *N* is the number of contributors under consideration, as set by the analyst in Step 1 of the STRmix™ mixture analysis. We advocate using the product rule since this is investigative work and the product rule improves run time and lowers the total bias. Individuals giving an *LR* above some user defined list management value, x, are investigated. x may be set quite high, use e.g. 1,000,000 to reduce the risk of adventitious hits. This decision may be usefully informed by $x \geq \frac{N}{HR}$ where *N* is the number of people on the database and *HR* is the current database hit rate. We would suggest exceeding the value of $\frac{N}{HR}$ by some margin, say *y*. The risk that a reported hit is false is about $1-\frac{xyHR}{xyHR+N}$ or less.

4.2.3 Database search module

STRmix™ allows the user to search a deconvoluted DNA profile against a database directly, without the need for deriving a single source component. Following a deconvolution, one of the resulting files, the genotype probability distribution (.txt file) which contains all the accepted, weighted genotype combinations proposed during the deconvolution, is compared to a database of known individuals.

The Database Search function can be used to provide investigative information, or as a quality assurance tool for comparison of complex mixtures where contamination is suspected [2]. The *LR*s that form part of the output are intended to be for investigative purposes only. Once an individual/s of interest has been identified through the database search a full STRmix™ analysis should be conducted for the evidence input file(s) against the reference sample of the individual and an appropriate evaluative *LR* calculated.

Step	Action
1	Within **Investigation**, select **Database Search** (refer Figure 4.1). The Database Search window will open (see Figure 4.4).
2	You may either drag in and drop a previously analysed run folder of choice or navigate to it using the **Browse** button. If you have selected **Investigate** at the end of an Interpretation or if you have selected a Previous Interpretation in the Investigation screen, this field will be pre-populated. To select a file via navigation, navigate to the folder, highlight the config.xml file (or Settings.ini) and select **Open** (see Figure 4.5). This returns you to the Database Search window.
3	Either drag and drop a Database of choice into the Previous Interpretation field or select **Browse** to open a file chooser. A database can also be set by default in Administration>STRmix Defaults (refer Section 7). Navigate to the database of choice and select **Open**.
4	Enter the list management value into the **Minimum LR** window or leave as zero if you would like to return all results. Only profiles with *LRs* above the value entered here will be returned.
5	Select the appropriate allele frequency file from the **Population for Search** drop down box.
6	Select **Standard** (from the Type of Search drop-down menu) to undertake a standard database search or **Familial** to undertake a familial search of the selected database.
7	If running a familial search, enter the appropriate mutation rate (default = 0.001).
	If running a standard search, optionally include an F_{ST} value and select the Assign Sub-Source LR calculation.
8	The Search ID will take the form of Case name_Sample name_DBSearch. This is an editable field.
9	Select **Start** to start the search or **Cancel** to return to the Investigation window.
10	On completion of the search, select **Finish** to close the window.

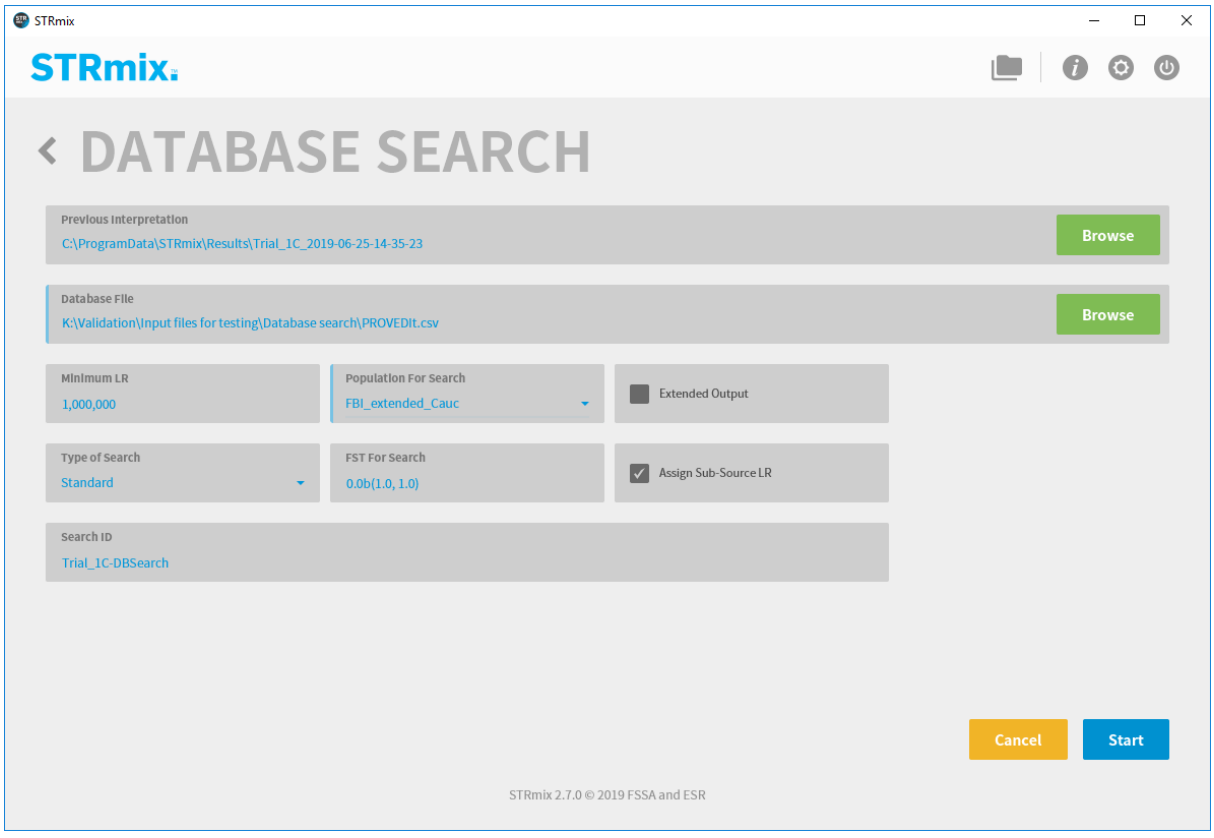


Figure 4.4 Database Search window

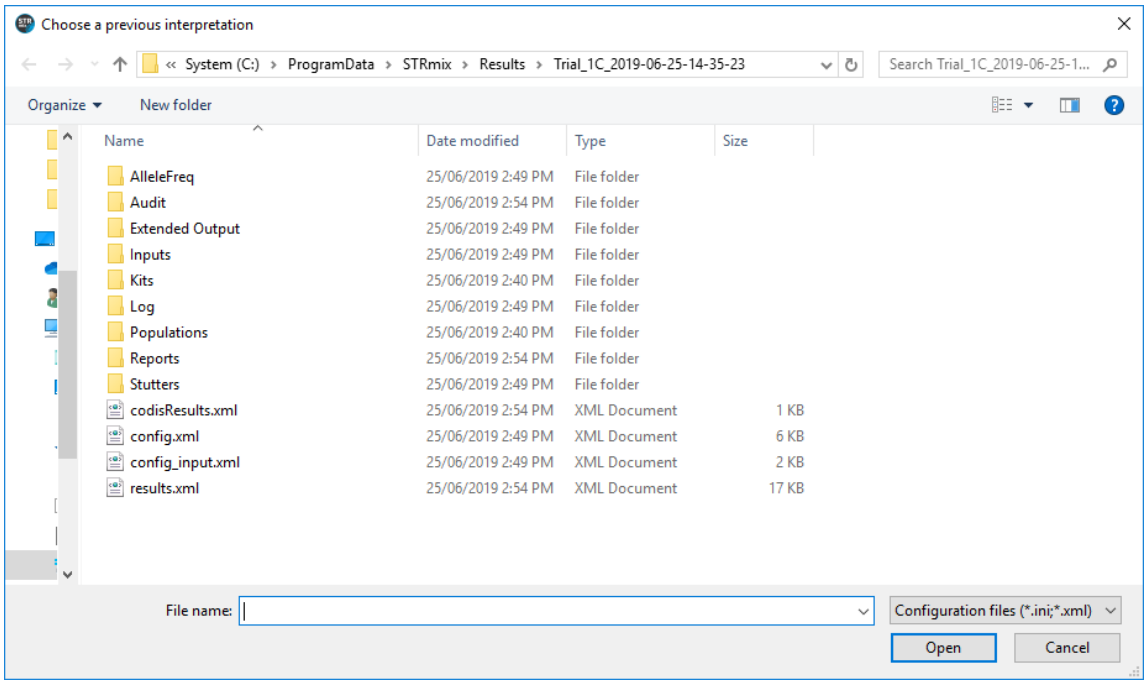
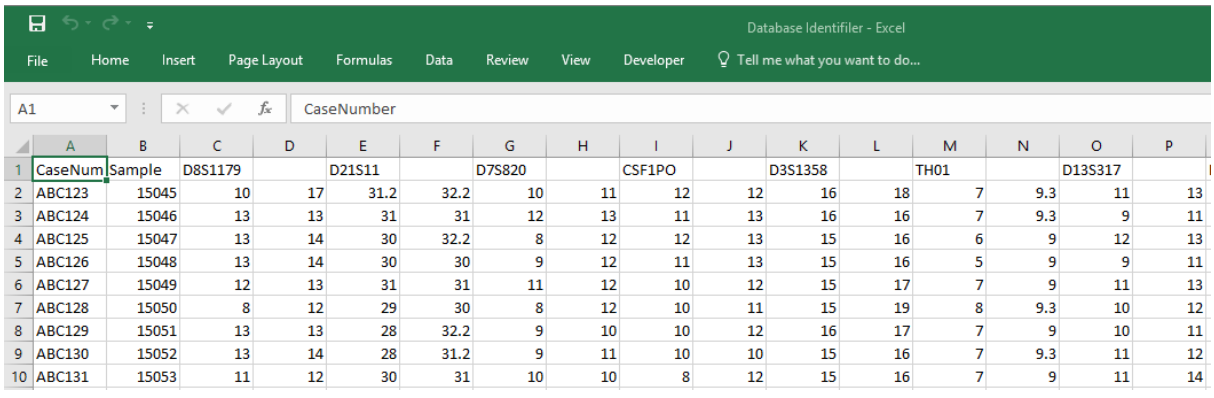


Figure 4.5 Database search file selection window

Notes

The database file must be a .csv type file in the following format (see Figure 4.6):

CaseNumber,Sample,Locus1Allele1,Locus1Allele2,…,LocusNAllele2



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	CaseNum	Sample	D8S1179		D21S11		D7S820		CSF1PO		D3S1358		TH01		D13S317	
2	ABC123	15045	10	17	31.2	32.2	10	11	12	12	16	18	7	9.3	11	13
3	ABC124	15046	13	13	31	31	12	13	11	13	16	16	7	9.3	9	11
4	ABC125	15047	13	14	30	32.2	8	12	12	13	15	16	6	9	12	13
5	ABC126	15048	13	14	30	30	9	12	11	13	15	16	5	9	9	11
6	ABC127	15049	12	13	31	31	11	12	10	12	15	17	7	9	11	13
7	ABC128	15050	8	12	29	30	8	12	10	11	15	19	8	9.3	10	12
8	ABC129	15051	13	13	28	32.2	9	10	10	12	16	17	7	9	10	11
9	ABC130	15052	13	14	28	31.2	9	11	10	10	15	16	7	9.3	11	12
10	ABC131	15053	11	12	30	31	10	10	8	12	15	16	7	9	11	14

Figure 4.6 Example of .csv file for database search

The first row of the database file is reserved for headers. Locus names must be identical to the names within the kit although the locus order does not matter. Each subsequent row contains a reference profile. The file must contain Amelogenin and all non-matching loci (such as Indels or Y STR markers) although these values are not used in the search. If there are missing loci in the database profiles (for example if comparing larger STR multiplex results with database profiles generated from smaller multiplexes) then label the loci headers, but leave the reference cells blank where allele information is missing.

If information at a locus is present, it must be complete, ie two allele values entered per locus. Leaving one cell blank will result in that profile failing the quality review. Any reference profiles that contain unrecognised data will be flagged after the quality review of the database file and a summary is provided in the report. An example of one profile not meeting quality requirements can be seen in the 'Data Validation' section in the report below (see Figure 4.7). Upon investigation, this reference profile was found to have incomplete data at one locus (ie information entered for one allele, not both).

The minimum *LR* value is used to filter out of the Results file all comparisons that lead to an *LR* below the specified value. Setting this to 0 will return an *LR* for all database entries.

The Population is used to select an allele frequency database for use in the *LR* calculations for a standard database match. All database comparisons are calculated by considering the database individual (POI) in an '*N*' person mixture by:

H_p = POI + *N*-1 unknowns

H_d = *N* unknowns.

A mutation rate is required for a familial search. The mutation rate takes into account the frequency of meiotic mutations and is used within the *LR* calculation for parent/child relationships at loci where there is an exclusion. There is no need to use a precise mutation rate for each locus if the search is being used in an investigative sense. The value 0.001 is an adequate approximation for autosomal STR loci.

Database search defaults including a default database, a default population for search, mutation rate for familial searches etc… can be set in the STRmix™ Defaults within the Administration module (ADMINISTRATION).

A database search can be set to run automatically upon completion of a deconvolution by setting this as the default option within STRmix™ Defaults (section 7).

4.2.4 An explanation of the database search results

When the Database Search is complete, the results are automatically presented on the screen. Results are automatically saved in a results directory named in part 'DBSearch'.

Select **Open Database Search Results Directory**  to open the appropriate results folder.

The results summarise the input files, deconvolution, previous interpretation details, database search run settings, and list of individuals (and the results of the gender loci if present within the database file) whose comparison to the genotype PDF has yielded an *LR* above the defined cut-off value. In the example in Figure 4.7 the minimum *LR* for list management was set to 1,000,000.

Profiles may be sorted by *LR* from highest to lowest (refer Section 7.5).

Database Search Report

DETAILS

STRMIX VERSION:	STRmix 2.7.0
USER:	jmorawitz
RUN DATE:	27 Jun 2019 14:45:22
REPORT RUN:	27 Jun 2019 15:37:41

DATABASE SEARCH SETUP

Interpretation chosen	C:\ProgramData\STRmix\Results\Trial_1C_2019-06-25-14-35-23
Profiling Kit	GlobalFiler_3500_GS
Sample file	Example 1C evidence_EV.csv
Database searched	K:\Validation\Input files for testing\Database search\PROVEDIt partial locus.csv
Number of Individuals in database	250
Minimum LR	1000000
Population	FBI_extended_AfAm
Type of search	Standard
FST	0.0b(1.0, 1.0)
Assign sub-source LR	Y
Extended output	N
Search ID	Trial_1C-DBSearch

PREVIOUS INTERPRETATION DETAILS

Interpretation Folder	C:\ProgramData\STRmix\Results\Trial_1C_2019-06-25-14-35-23
Interpretation Run Date	25 Jun 2019 14:35:23
Case Number	Trial
Sample ID	1C
Case Notes	
Interpretation Seed	996488
Number Of Contributors	2

DATA VALIDATION

Number of Invalid ref files	1
Invalid ref file	RD14-0003_K1

LR RESULTS

CASE NUMBER	SAMPLE	GENDER	LR
RD14-0003	K32	X,Y	1.5171E26

Figure 4.7 Database Search Results (standard search example)

4.3 Hd True Tester

Select **HD True Tester** from the **Investigation** screen (see Figure 4.1) to open the H_d True Tester window (see Figure 4.8). The H_d True Tester tool allows the user to perform large *in-silico* specificity tests on a profile by profile basis.

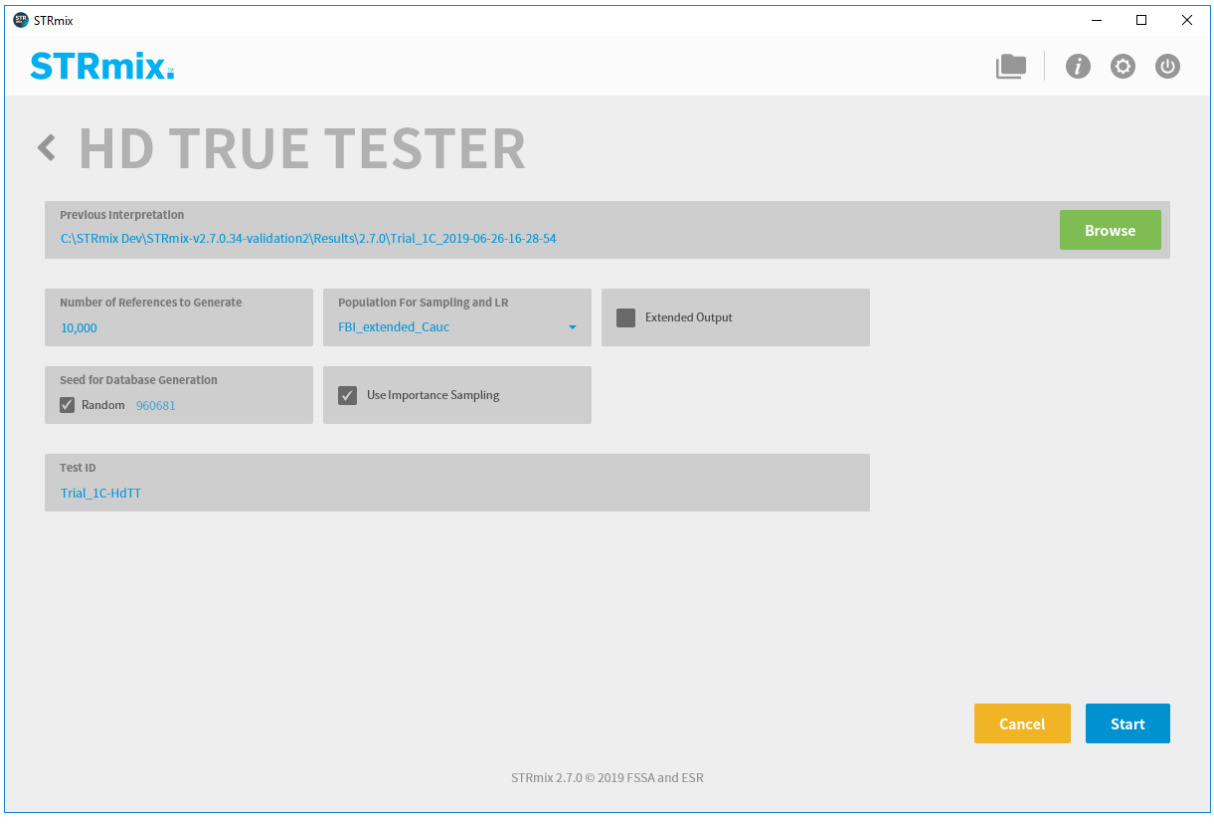


Figure 4.8 Hd True Tester window

Drag and drop the run folder of interest (if not undertaken in the Investigations window) into the Previous Interpretation field (see Figure 4.2) or navigate to the config.xml (or Settings.ini) file of the deconvolution you wish to add using the **Browse** button.

Select the population for the database generation and *LR* calculation. Edit the number of references to generate if required. Choose a starting seed if required (refer 3.3.1) and select **Start** to start the analysis. Results will be saved as Case name (of deconvolution)_Sample name-HdTT.

An *LR* will be calculated for each reference sample generated. Profiles are generated by sampling from the accepted weighted genotypes from the deconvolution (if the Use Importance Sampling tickbox is selected). For dropped alleles (-1 in the raw report or Q alleles in the pdf report) the allele is sampled from the selected population. The propositions considered are:

H_p: any assumed individuals + randomly generated individual within database + unknowns up to the NOC

H_d: any assumed individuals + unknowns up to the NOC.

Refer to the STRmix™ User's Manual for further details.

4.3.1 Hd True Tester results

When the H_d true test is complete, the results are automatically presented on the screen. Results are automatically saved in a results directory named in part 'HdTT'. Select **Open Hd True Tester Results Directory** to open the appropriate results folder.

4.4 Mix to Mix

The Mix to Mix function within STRmix™ compares each pair of contributors between two mixtures and calculates an LR for each pairwise comparison. This is based on the method Slooten [3] described to calculate an LR for common donors to two profiles, M and M', where there is no longer the requirement for one of the profiles to be single-source (termed mixture to mixture matching due to its extension of standard matching) [4-6].

The hypotheses are:

H_1: D_1=D_1' and all other donors of the mixture are unrelated

H_2: All donors of both mixtures are unrelated

Where:

D_1 is contributor 1, mixture 1 and

D_1' is contributor 1, mixture 2

Mix to Mix matching also calculates the average of the LRs, this is the equivalent of:

H_1: The two mixtures share one common contributor

H_2: All contributors to both mixtures are unrelated

4.4.1 The Mix to Mix module

Select **Mix to Mix** from the Investigation screen (see Figure 4.1 STRmix™ Investigation window to open the Mix to Mix window (see Figure 4.9). If a Previous Investigation is already loaded in the Investigation screen (eg if the **Investigate** button was selected at the end of an interpretation), this interpretation will be added to the Left Interpretations list on the left side of the screen.

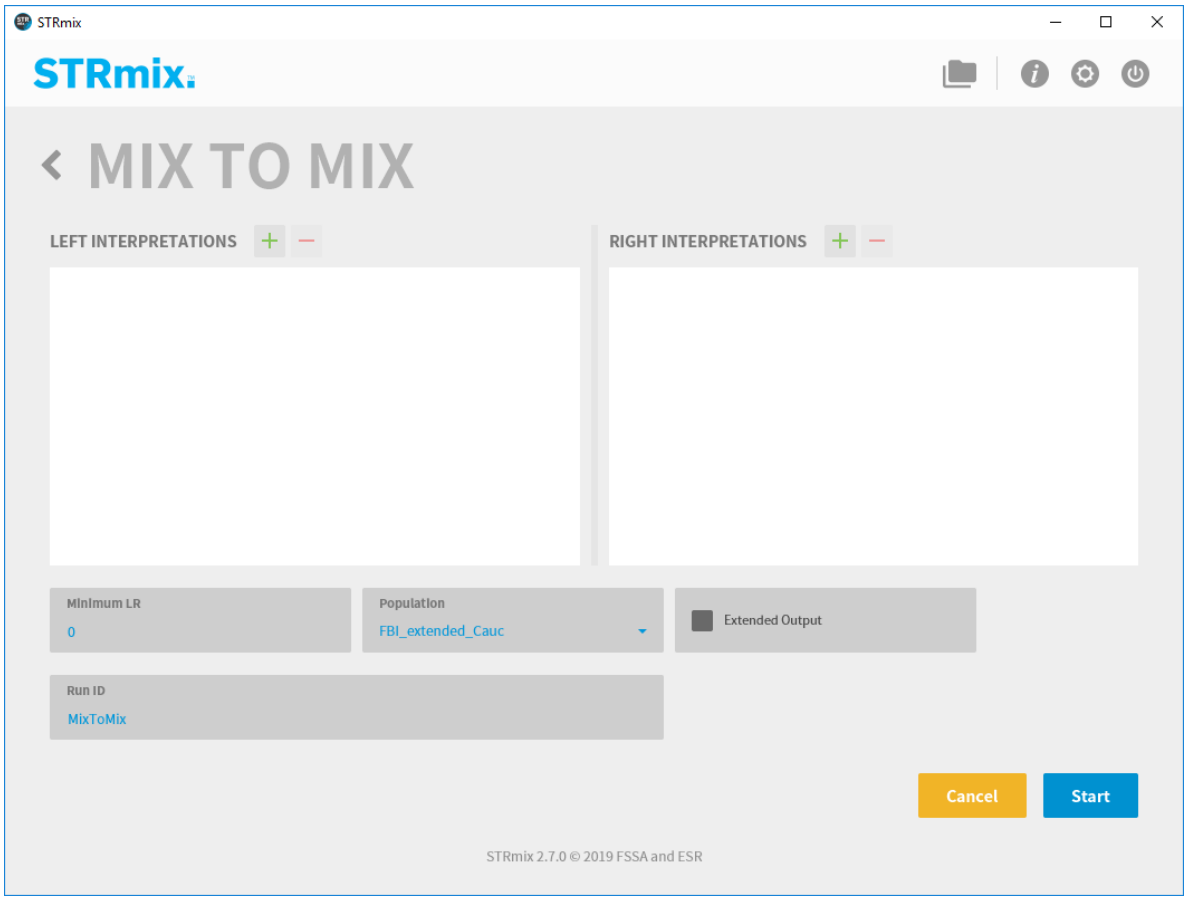


Figure 4.9 Mix to Mix window

From this screen previous interpretations can be added to the left hand side and right hand side either by dragging and dropping or by using the **Add previous interpretation** button ("**+**") to Browse to find the relevant config.xml (or Settings.ini). There is no requirement for the number of contributors to be the same between the mixtures.

Interpretations added to the left side of the screen will be compared with interpretations added to the right side of the screen, via each contributor position in turn (see Figure 4.10). Note that deconvolutions within the same panel will not be compared with each other.

For example, if the first interpretation on the left is a three person mixture and the first interpretation on the right is a two person mixture, the pairwise comparisons would be as follows:

Mix1 contributor1 to Mix2 contributor1

Mix1 contributor1 to Mix2 contributor2

Mix1 contributor2 to Mix2 contributor1

Mix1 contributor2 to Mix2 contributor2

Mix1 contributor3 to Mix2 contributor1

Mix1 contributor3 to Mix2 contributor2

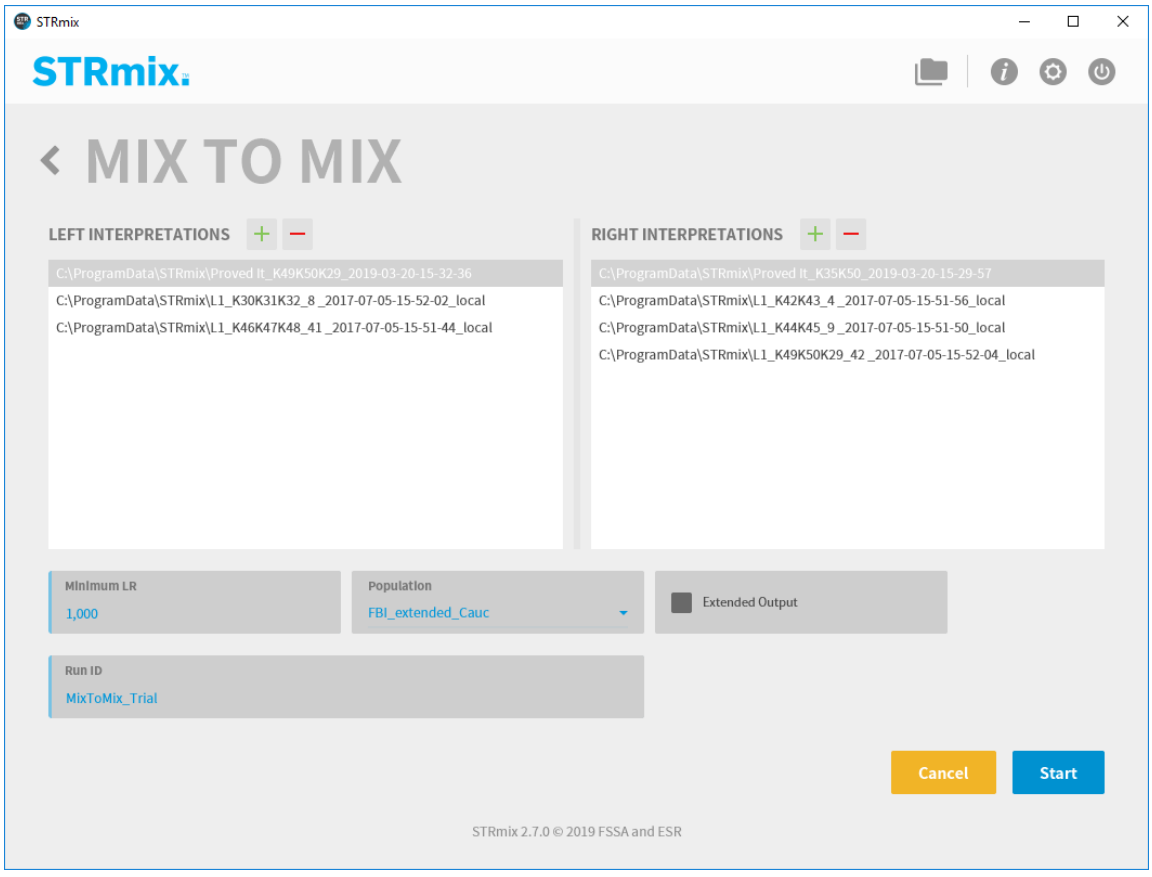


Figure 4.10 Mix to Mix window with interpretations added

Enter the list management value into the **Minimum LR** field or leave as zero if you would like to return all results. Only profiles with *LR*s above the value entered here will be returned.

Select the appropriate allele frequency file from the **Population**drop down box.

Edit the **Run ID** field if required and select **Start** to start the investigation or **Cancel** to cancel and return to the Investigation window.

4.4.2 An explanation of the Mix to Mix results

When the Mix to Mix investigation is complete, the results are automatically presented on the screen. Results are automatically saved in a results directory named in part 'MixtoMix'.

Select the folder icon for the Mix to Mix Results Directory to open the appropriate results folder.

The results summarise Mix to Mix setup and then the input files, sample details (including the number of contributors in each sample) and the *LR* results for each set of comparisons (see Figure 4.11). In the example, the minimum *LR* for list management was set to 1,000.

Mix to Mix Report

DETAILS

STRMIX VERSION:	STRmix 2.7.0
USER:	jmorawitz
RUN DATE:	12 Jul 2019 16:19:53
REPORT RUN:	12 Jul 2019 16:37:58

MIX TO MIX SETUP

Minimum LR	1,000
Population chosen for calculation	FBI_extended_Cauc
Number of interpretation comparisons performed	12
Number of interpretation comparisons meeting minimum LR threshold	2

LR RESULTS

COMPARISON (1 OF 2)

Interpretation 1	
Case number	Proved It
Sample name	K49K50K29
Number of contributors	3
File location	C:\ProgramData\STRmix\Proved It_K49K50K29_2019-03-20-15-32-36
Interpretation 2	
Case number	Proved It
Sample name	K35K50
Number of contributors	2
File location	C:\ProgramData\STRmix\Proved It_K35K50_2019-03-20-15-29-57

INTERP 1 NOC	INTERP 2 NOC	INTERP 1 COMPONENT	INTERP 2 COMPONENT	LR
3	2	1	1	1.3825E26
		1	2	1.6221E-5
		2	1	1.6991E23
		2	2	3.0931E-1
		3	1	3.6433E18
		3	2	1.4010E1
			AVERAGE LR	2.3070E25

Figure 4.11 Example of Mix to Mix results

4.5 Investigation Batch

The Investigation Batch tool allows batching of Database Searches and *LR* from Previous investigations on previous interpretations. Interpretations can be set up to run against one or more references or Databases under different conditions if required.

Select **Investigation Batch** from the Investigation screen (see Figure 4.1) to open the Investigation Batch window.

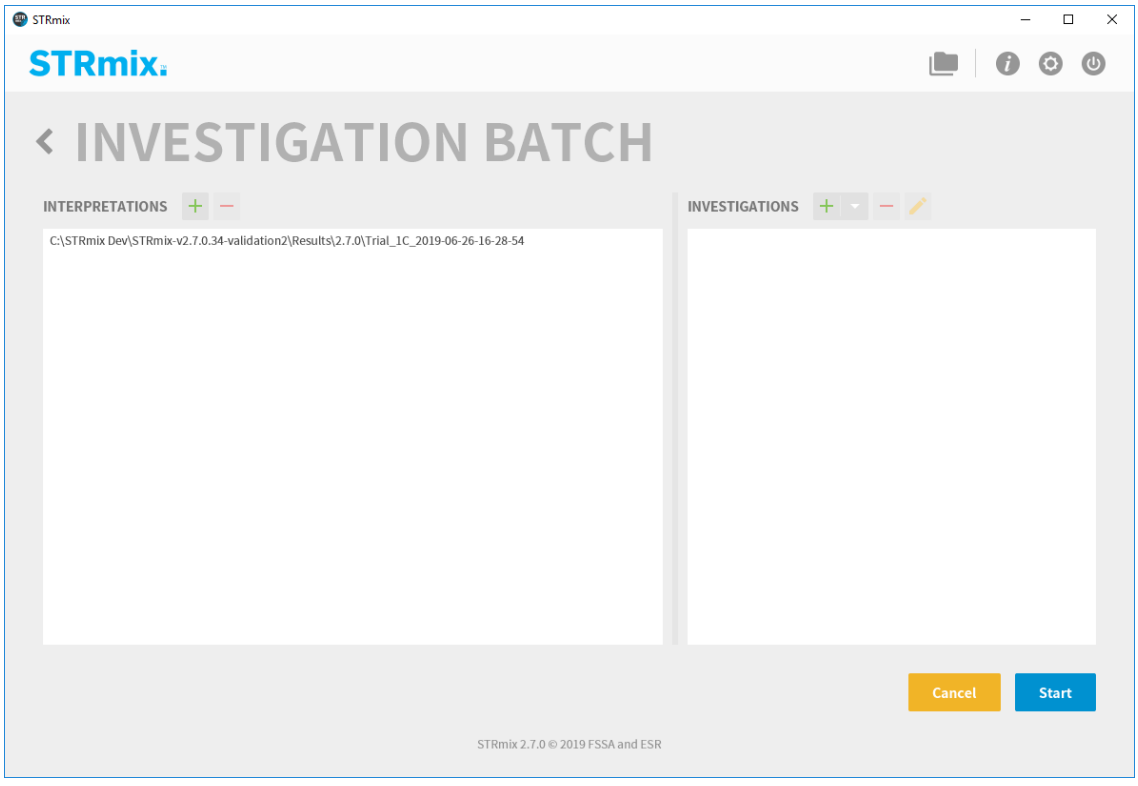


Figure 4.12 Investigation Batch window

From this screen previous interpretations can be added to the left hand side either by dragging and dropping or by using the **Add previous interpretation** button ("**+**") to Browse to find the relevant config.xml (or Settings.ini).

4.5.1 LR from Previous Batch (within Investigation Batch)

To set up an LR Batch, select **Add LR from Previous** from the drop down menu on the right hand side of the screen. Add reference inputs using the **Add Reference File** button ("**+**") or by dragging and dropping the config.xml or the run folder. The *LR* Batch tool allows the user to calculate multiple *LR*s from multiple reference inputs to a previously run deconvolution and vice versa. Note that the kits do not need to be the same for each deconvolution and the reference sample/s.

More than one LR from Previous Batch can be set up at a time (for example if different *LR* Settings are required). Each will be appended with a Run ID (see Figure 4.13). Populations, F_{ST}, and other *LR* Settings can be changed as for any other *LR* calculation. Choose a starting seed if required (refer 3.3.1). Select **Save** to save the batch or **Cancel** to return to the Investigation Batch screen. Select **Start** to add the *LR* from Previous Batch(es) to Batch Mode, or **Cancel** to return to the Investigation window.

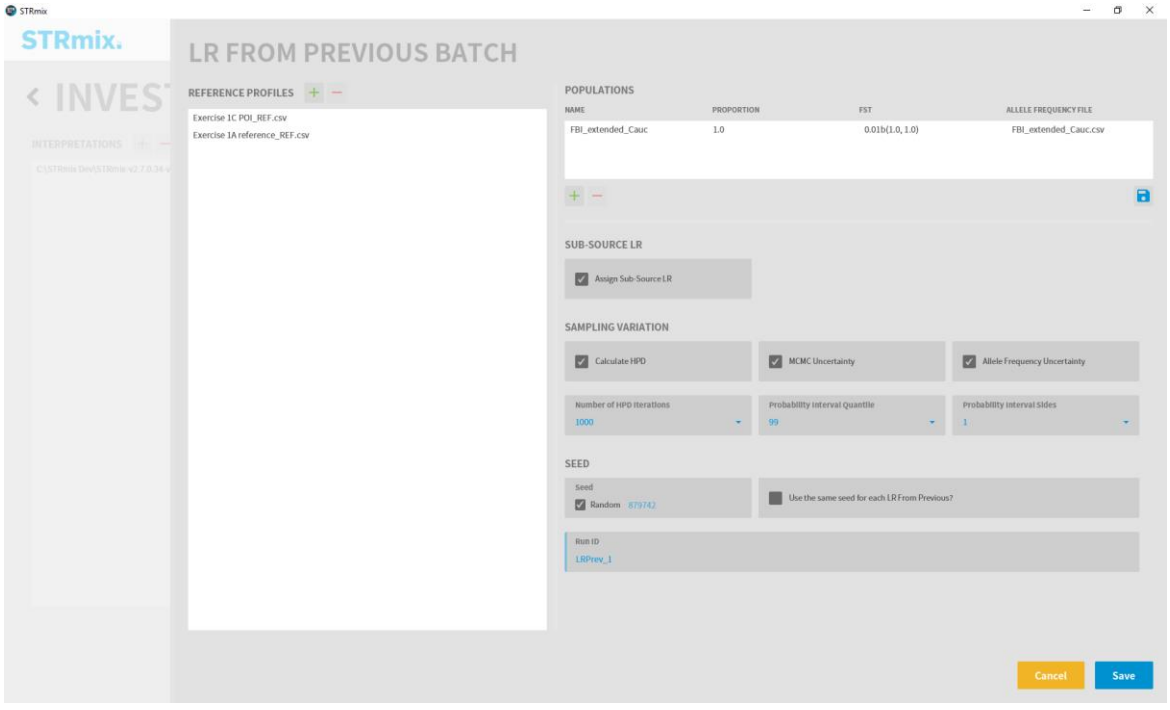
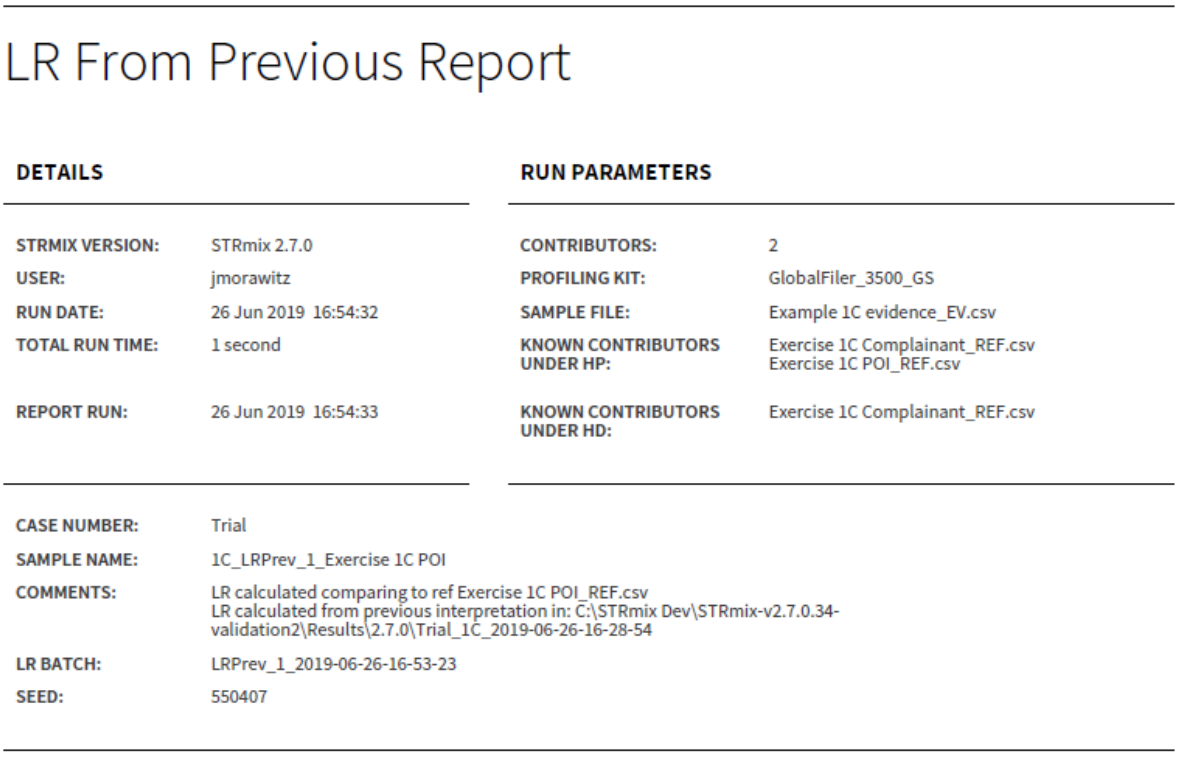


Figure 4.13 LR from Previous Batch window

Results will be saved as Case name (of deconvolution)_Sample name_LRPrev_X_Reference profile compared to within the Results folder. An *LR* will be calculated for each reference sample entered to each deconvolution entered.

Results for each deconvolution comparison with reference will include Comments detailing the samples being compared (See Figure 4.14).



LR From Previous Report

DETAILS

STRMIX VERSION:	STRmix 2.7.0
USER:	jmorawitz
RUN DATE:	26 Jun 2019 16:54:32
TOTAL RUN TIME:	1 second
REPORT RUN:	26 Jun 2019 16:54:33

RUN PARAMETERS

CONTRIBUTORS:	2
PROFILING KIT:	GlobalFiler_3500_GS
SAMPLE FILE:	Example 1C evidence_EV.csv
KNOWN CONTRIBUTORS UNDER HP:	Exercise 1C Complainant_REF.csv Exercise 1C POI_REF.csv
KNOWN CONTRIBUTORS UNDER HD:	Exercise 1C Complainant_REF.csv

CASE NUMBER:	Trial
SAMPLE NAME:	1C_LRPrev_1_Exercise 1C POI
COMMENTS:	LR calculated comparing to ref Exercise 1C POI_REF.csv LR calculated from previous interpretation in: C:\STRmix Dev\STRmix-v2.7.0.34-validation2\Results\2.7.0\Trial_1C_2019-06-26-16-28-54
LR BATCH:	LRPrev_1_2019-06-26-16-53-23
SEED:	550407

Figure 4.14 Case details excerpt from LR Batch Results

4.5.2 Database Search Batch (within Investigation Batch)

The Investigation Batch tool allows any number of Database searches (and/or LR from Previous runs) to be set up against the same previous interpretation or set of previous interpretations with different search conditions (including the option to add a different Database File for each batch).

In the Investigation Batch window (see Figure 4.12), to set up a Database Search batch, select **Add Database Search** from the drop down menu on the right hand side of the screen. Add a Database file using the **Browse** button or by dragging and dropping the database .csv file into the Database File field. Select the appropriate settings for the Database Search and then **Save** or **Cancel** (see Figure 4.15)**.** The Database Search Batch tool allows the user to carry out multiple Database searches to multiple previously run deconvolutions.

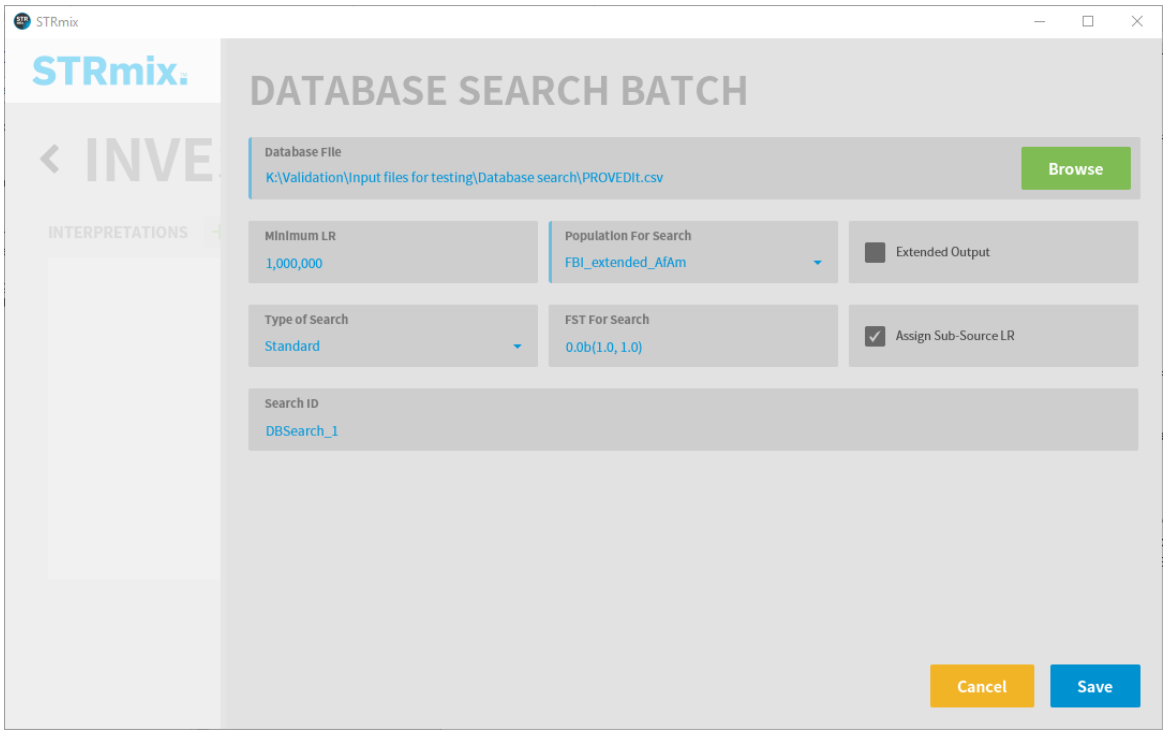


Figure 4.15 Database Search Batch window.

More than one Database Search Batch can be set up at a time, for example if different Database Search Settings or different databases are required. Each will be appended with a Search ID (see Figure 4.15). The type of search and other search settings can be changed as for any other Database Search calculation. Select **Save** to save the batch or **Cancel** to return to the Investigation Batch screen. Select **Start** to add the Database Search Batch(es) to Batch Mode or **Cancel** to return to the Investigation window.

Notes

The **Remove selected interpretation** or **reference** buttons ("**-**") can be used to remove a deconvolution (Figure 4.12 and Figure 4.13) (or reference sample if setting up an LR from Previous batch) from the relevant window after selecting the sample.

The **Remove selected investigation** button ("**-**") can be used to remove an investigation from the Investigations window after selecting the investigation to remove (Figure 4.12).

You can choose to add either a deconvolution that has had an *LR* calculated for it previously using a reference sample or a deconvolution only that has been run with no reference samples. *LR* from Previous Batch and Database Search Batch use only the deconvoluted results from the previously run analysis.

Database Search batches can be set up independently from or at the same time as an *LR* from Previous Batch. For example, a Staff Elimination Database search could be set up at the same time as an *LR* from Previous batch with a number of case specific references all to the same deconvolution(s).

The Investigation Batch function does not allow the use of the User Selected priors method on a varNOC interpretation; however, it does allow methods of Stratify or MLE. User Selected numbers of contributors for H_p and H_d for a varNOC analysis can be set outside of

Investigation Batch, either during the Interpretation or in a standard *LR* from Previous or a standard Database Search (see section 3.3.6.3).

4.6 Top Down Approach

The Top Down Approach to DNA mixture interpretation allows a user to set the number of major contributors to a mixed DNA profile they are interested in and obtain an *LR* only for these. For example in a five person mixture with two major contributors and three low level trace contributors, the user may be interested in determining a likelihood ratio for the two major contributors. This method could be used to screen complex DNA mixtures to determine their evidential value prior to carrying out a full deconvolution on them.

Select **Top Down Approach** from the Investigation screen (see Figure 4.1) to open the Top Down Approach set-up window (see Figure 4.16). Note that if Database search defaults have not previously been set within the Administration area (see section 7.1), then a warning message will appear prompting the user to first link a database to search and a default population for the Top Down Database Search.

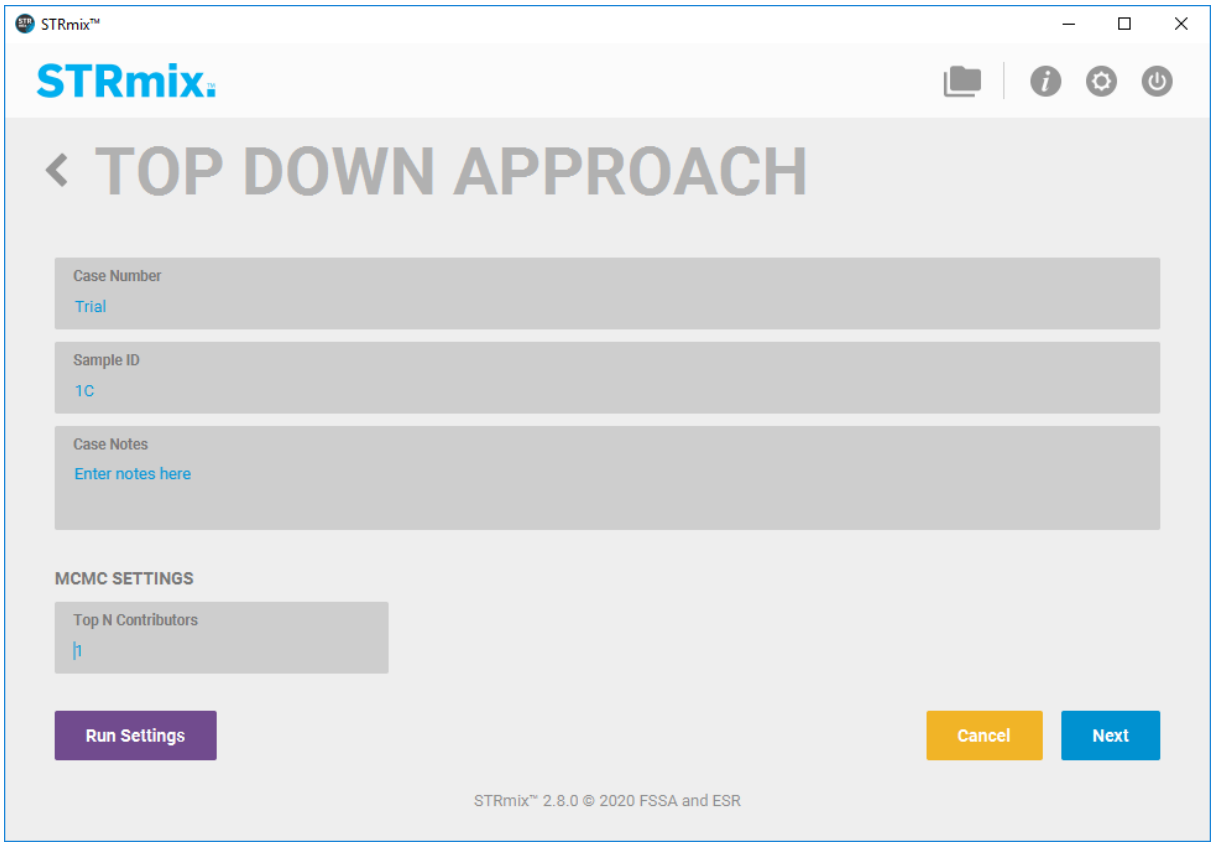
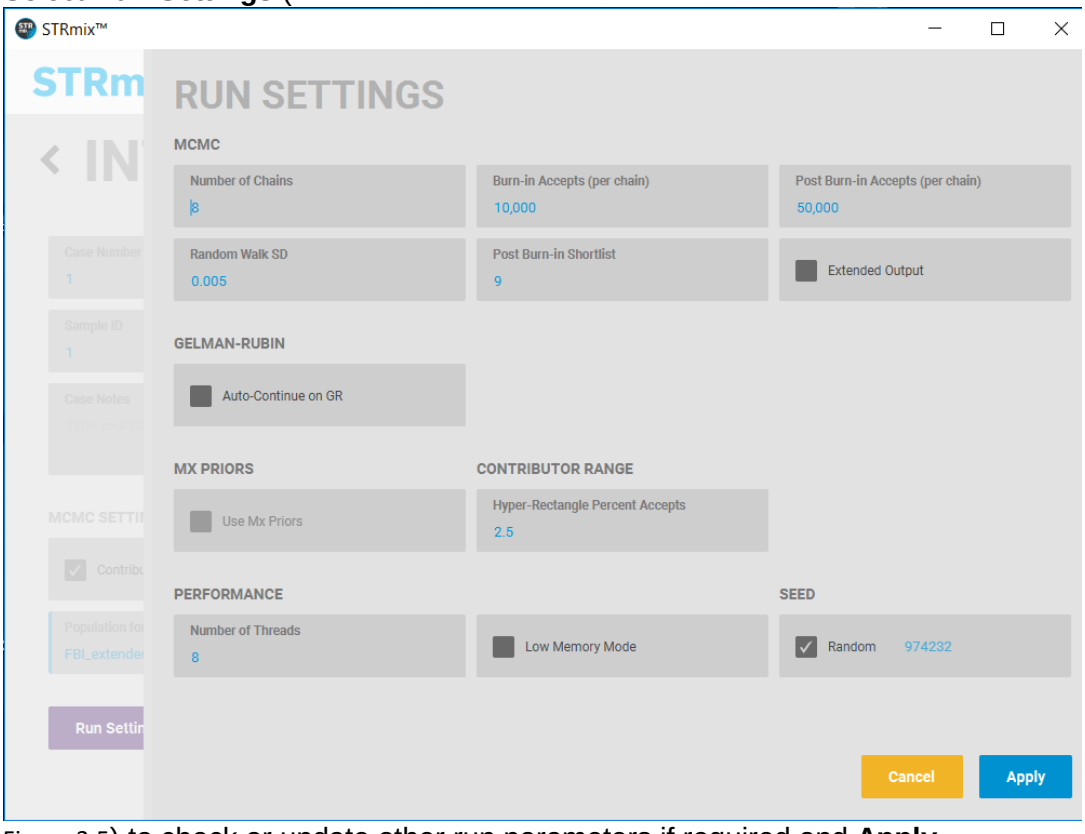


Figure 4.16 Top Down Approach set-up window

Step	Action
1	Ensure that a default Database and population are set in Administration (See Section 7.1)
2	Complete case number and sample ID information and optionally enter case notes

3	Select the top N number of contributors you are interested in searching. Note that there is no need to specify the number of contributors in the mixed DNA profile here
4	Select **Run Settings** ( Figure 3.5) to check or update other run parameters if required and **Apply**
5	Select **Next** to proceed to adding profile data or **Cancel** to return to the Investigation screen
6	Select a Profiling Kit from the drop-down menu (refer Figure 4.17)
7	Select **Kit Settings** to check or update the kit settings (eg ignore loci) if required and **Apply**
8	Add an evidence to the **Evidence Profile Data** field either by dragging and dropping the .txt or .csv file in the field or using the '**+**' button to Browse to the location of the input file to select it
9	Add any reference profiles to be conditioned on (or assumed) into the **Reference Profile Data (Hd known only)** field
10	Ensure the Perform Database Search tickbox is ticked (this will be ticked by default) and select **Start** to start the search, **Back** to return to the Top Down Approach set-up screen or **Cancel** to cancel the operation.

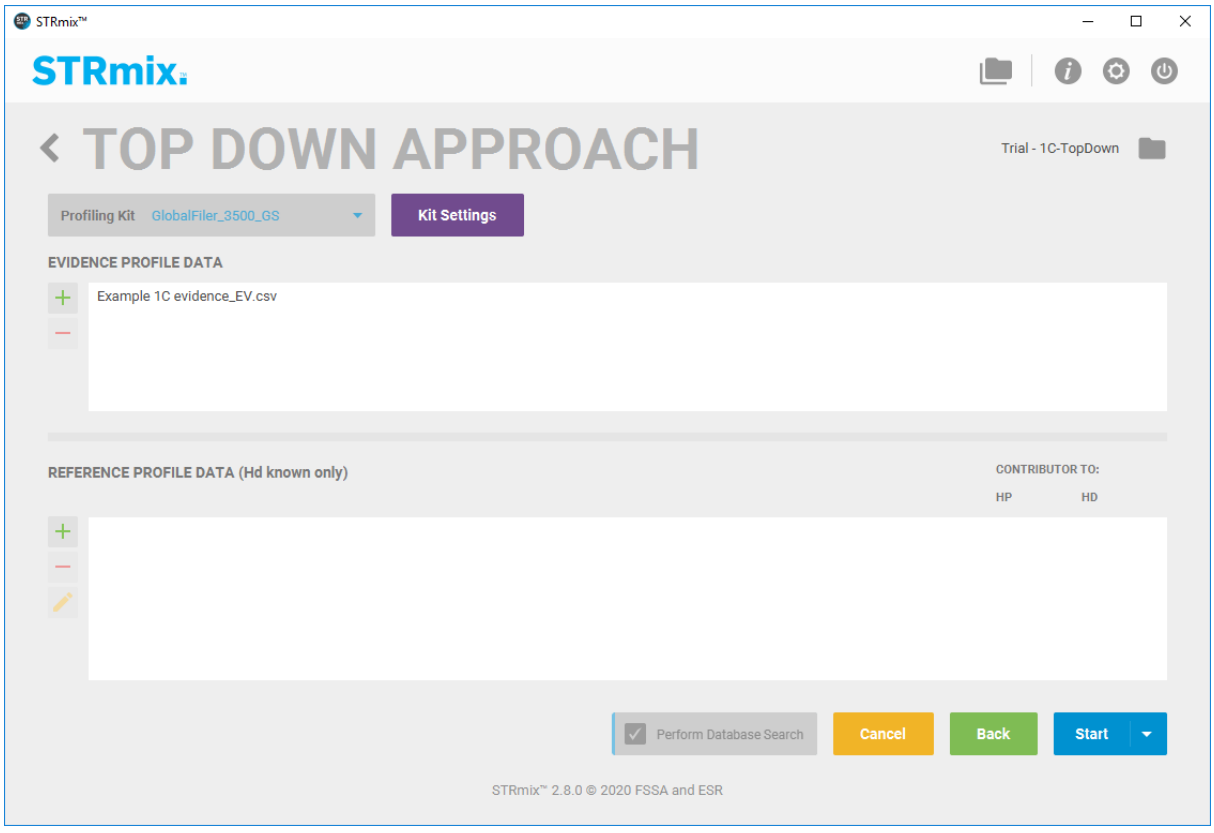


Figure 4.17 Top Down Approach add profile data window

Notes:

- The Top Down Approach utilises the standard Database search and cannot be run using a Familial search.
- Mx Priors cannot be used in the Top Down Approach
- Multi-kit analysis cannot be carried out in the Top Down Approach
- Replicate input files cannot be used in the Top Down Approach
- Running Extended Outputs when using the Top Down Approach will result in an extended output for each of the 'Steps' interpretation run folders.

4.6.1 An explanation of the Top Down results

When the Top Down Search is complete, the results are automatically presented on the screen. Results are automatically saved in a results directory named in part 'TopDown'.

Select **Open Top Down Approach Results Directory** to open the appropriate results folder.

The report and results files summarise the evidence input file, top down run settings, audit hashes and list of individuals (and the results of the gender loci if present within the database file) whose comparison to the genotype PDF has yielded an *LR* above the defined cut-off value. In the example in Figure 4.7, the minimum *LR* for list management was set to 1,000,000.

Profiles may be sorted by *LR* from highest to lowest (refer Section 7.5).

Top Down Approach Report

DETAILS

STRMIX VERSION:	STRmix 2.8.0
USER:	jmorawitz
RUN DATE:	18 Sep 2020 15:33:24
TOTAL RUN TIME:	8 minutes, 0 seconds
REPORT RUN:	18 Sep 2020 15:48:21

RUN PARAMETERS

TOP N CONTRIBUTORS:	2
PROFILING KIT:	GlobalFiler_3500_GS
SAMPLE FILE:	K30K31K32_52_EV.csv

CASE NUMBER:	Trial
SAMPLE NAME:	3 person mixture
COMMENTS:	
SEED:	576374

DATA VALIDATION

All data passed validation

LR RESULTS

CASE NUMBER	SAMPLE	GENDER	MAX LR
RD14-0003	K31	X,Y	1.0525E20
RD14-0003	K32	X,Y	1.4508E10

Figure 4.18 First page of a Top Down report for a three person mixture where N was set to 2

The results folder for a Top Down search contains the same files as a standard Database Search with the addition of a Steps folder (see Figure 4.19). The Steps folder contains the deconvolution and database search results for each step taken in the Top Down search at each different AT (see Figure 4.20).

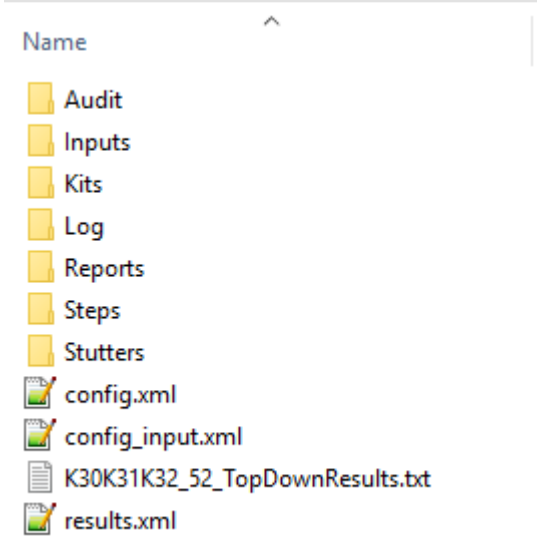


Figure 4.19 Example of a results folder from a Top Down search

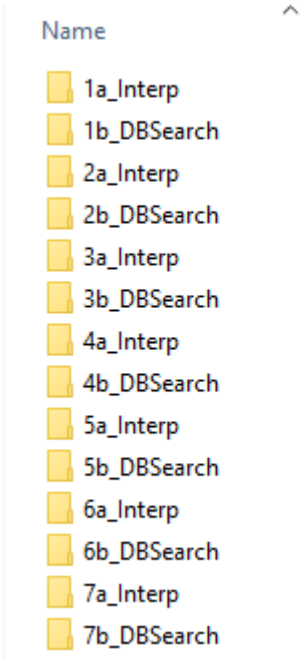


Figure 4.20 Example of part of the Steps folder from a Top Down search

5. BATCH MODE

Within Batch Mode, a number of different STRmix™ analyses can be set up and queued to run sequentially. There is no need for the interpretation to use the same kits, settings or numbers of contributors and one batch can contain variously Interpretations, Database searches, H_d True tests, Model Maker analyses, *LR* from Previous calculations etc. Each of the analyses are set up in turn and details are saved in individual folders as per standard STRmix™ analysis.

Select **Batch Mode** from the start-up screen (see Figure 3.2). The Batch Mode window will open (see Figure 5.1). Select **Add to Batch** to open the Interpretation window. Alternatively, select a different type of analysis from the dropdown list next to **Add to Batch**.

Complete the analysis set up. On completion, select **Queue** to return to the Batch Mode window. The case details will be entered into the Calculations in Batch window. Select **Add to Batch** to enter the next calculation. Repeat for as many samples as needed.

To remove a sample from the Batch Mode, highlight the sample in the Calculations in Batch window and select **Remove**.

Select **Change** to change the location where the results folders will be saved. Results from analyses added after changing the location will saved to the new location.

Select the STRmix™ logo or Back arrow (<) to return to the main directory window. The batch details will be saved until you return, even upon exiting STRmix™.

Enable **Run Replicates?** and select the number of replicates to run and the initial seed to set if required. This will replicate each sample in the batch the chosen number of times with a new start seed for each replicate. Note that this function is primarily for testing and

validation, and STRmix™ does not advocate the repeat interpretation of crime profiles for casework analysis.

Select **Start** to start the batch run. Selecting **Stop** at any time will stop the current interpretation and return the user to the Batch Mode window. The batch details are saved and the batch may be started at a later date. Once the batch has started the progress may be tracked for each case as with a standard STRmix™ analysis.

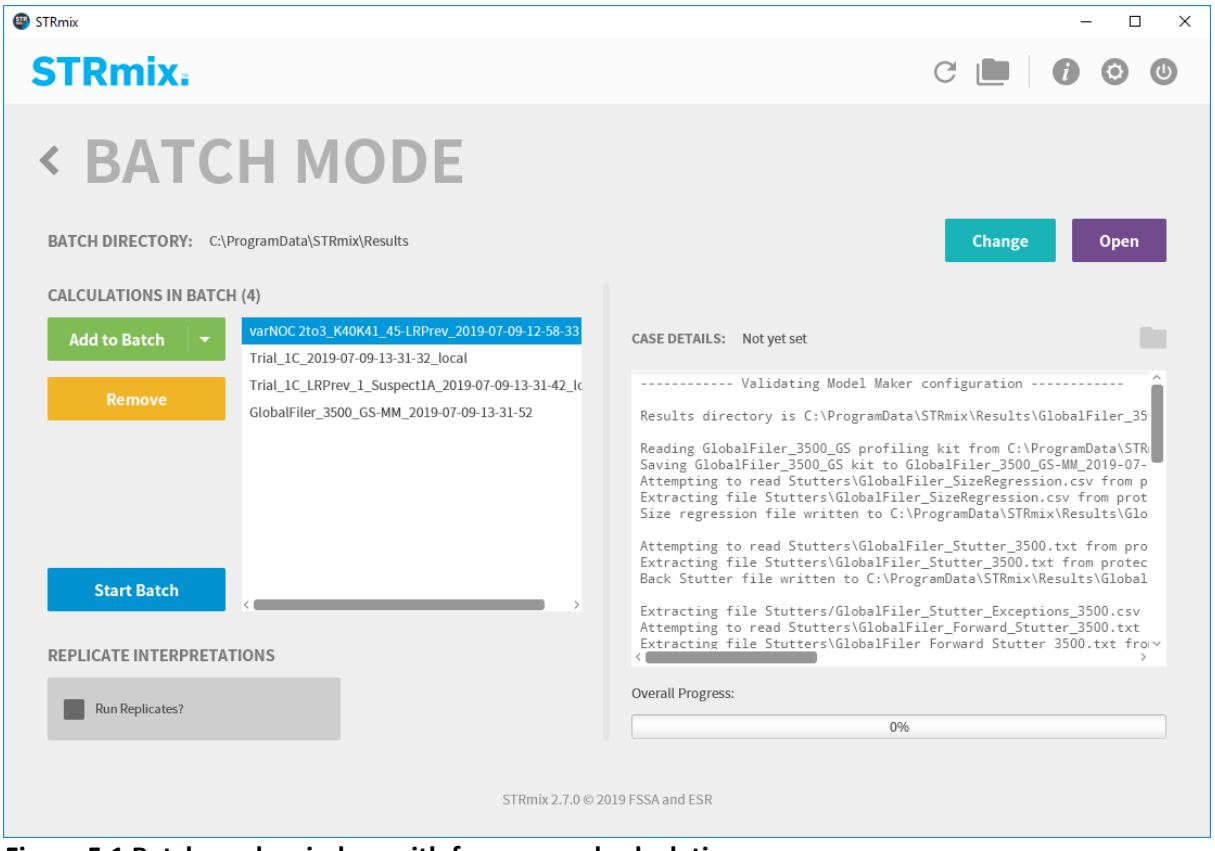


Figure 5.1 Batch mode window with four queued calculations

Notes

A batch mode file (.batch_mode) will be saved within each run folder created in batch mode. This file informs STRmix™ that this folder is to be run in a batch.

The STRmix™ screen can be minimised whilst running in Batch Mode to allow use of other computer functionality. There may be a noticeable impact on computer memory and speed, however.

The run folders for samples added in Batch Mode and waiting to be run will appear in the STRmix™ directory Results folder with a .batch_mode file inside each folder. When each run is completed the .batch_mode file is removed from the run folder. The analyses will execute in the order they were added.

The **Add to Batch** button will automatically open the last type of Interpretation used in Batch Mode during the STRmix session.

6. REPORTS

6.1 Retrospective Reports

Select **Reports** from the Main Menu screen (refer Figure 3.2) to open the Reports window. From this screen **Browse** or drag and drop to add a Previous Interpretation and select **Run** to open the Reports window (see Figure 6.1) or **Cancel** to return to the Main Menu. The Reports tool allows the user to generate tailored reports for completed STRmix™ analyses. Retrospective reports can be run on STRmix™ analyses generated in versions 2.5 through to 2.7.

6.1.1 Summary reports

Summary reports for deconvolutions, *LR* calculations, database and familial search results, H_d True Tester, and Model Maker results can be run retrospectively. Drag and drop the run folder of interest into the Previous Calculation window (see Figure 6.1) or navigate to the config.xml file of the analysis you wish to add using the **Browse** button. Where appropriate, select components to be included in the report by checking them on or off. Select the order that the components should appear in the report by dragging and dropping them in the list order. Components can also be forced to start on a new page by checking the **New Page** option.

Certain components have additional configurability, for example by highlighting the **LR Summary** component, the option to include or exclude the Unified and Stratified *LR*s is available (see Figure 6.1).

Only components relevant to the type of analysis that was run in STRmix™ will be available for configuration within the report.

The location that the results are to be saved can be changed by selecting the **Browse** button within the Output File field. This field also allows the file name to be configured.

Other formatting options for the reports can also be changed under **General Settings**. Such as the size of the report between A4 and US Letter, date formatting, a laboratory's logo on the report, the size of the bottom margin, whether or not to automatically open the report on completion, and whether or not to include audit hashes on the pdf report (see Figure 6.2).

Select Run to generate the report. Select Cancel to return to the Main menu.

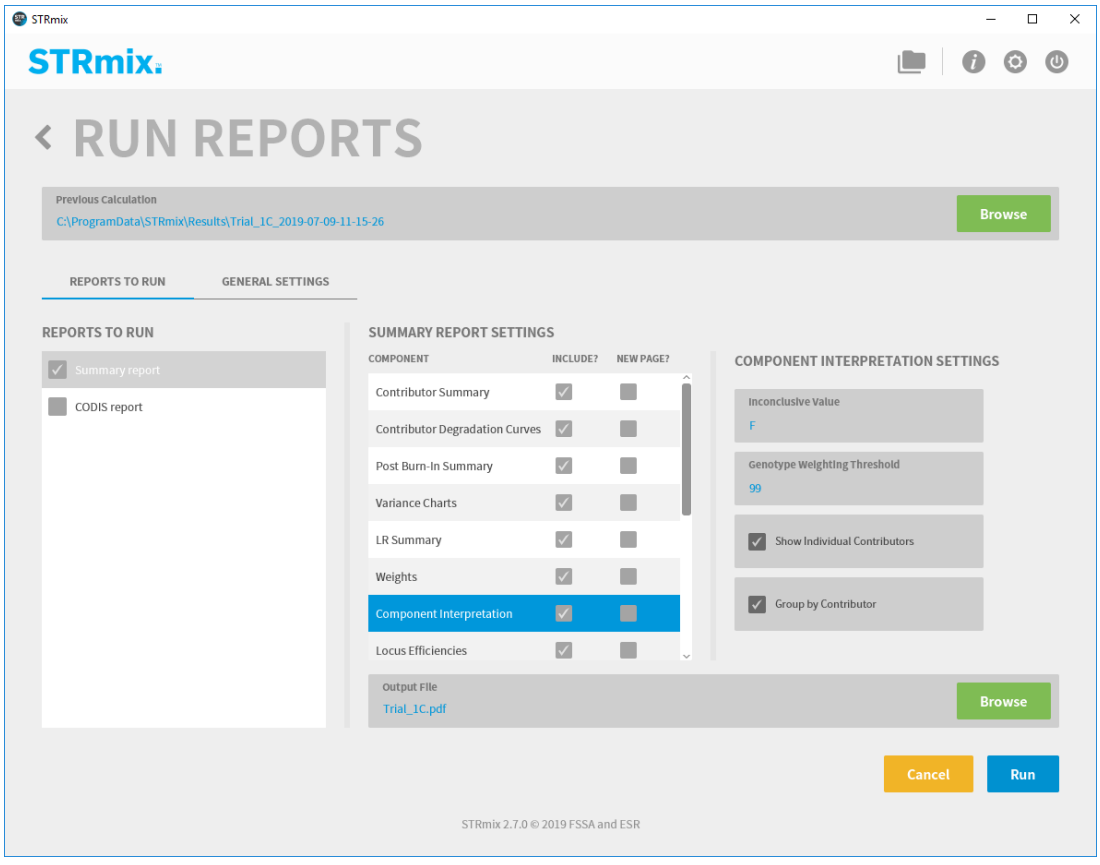


Figure 6.1 Reports to run tab

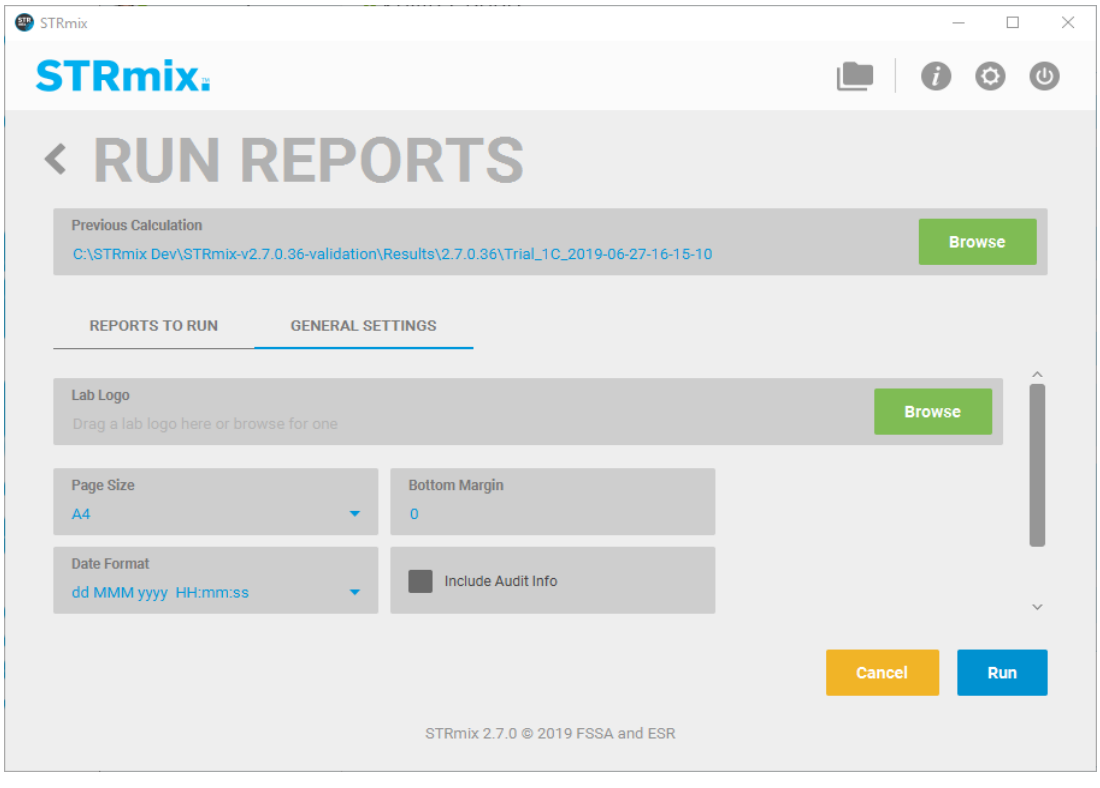


Figure 6.2 General settings tab

7. ADMINISTRATION

Select **Administration** from the Main Menu screen to open the Administration window. If enabled, you will be prompted to enter the Administrator password now (see Figure 7.1)

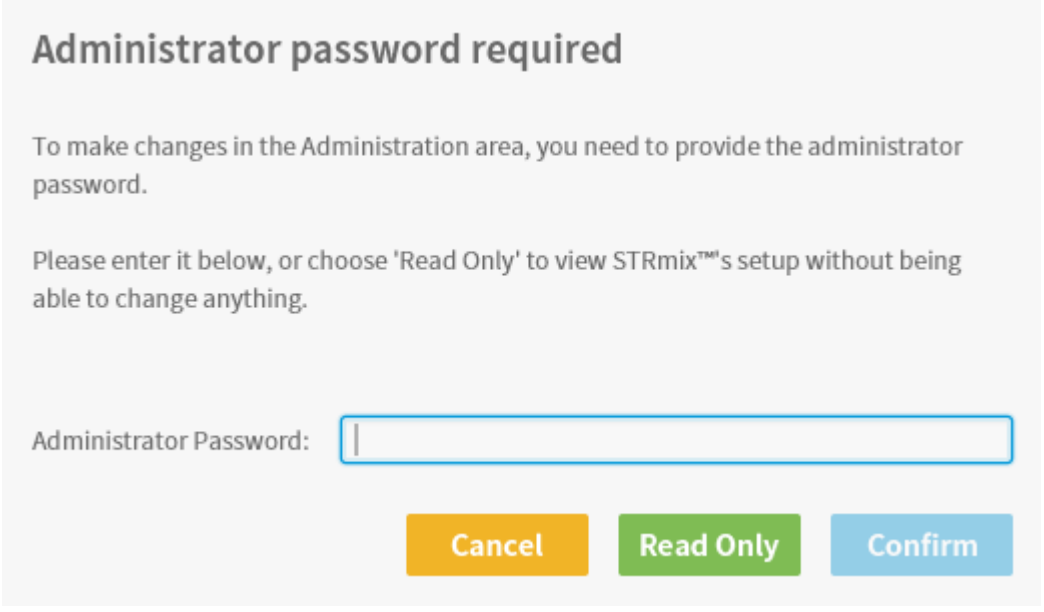


Figure 7.1 Password entry window

When you have entered the correct password, the **Confirm** button will be activated and you will be able to use this button to proceed to the Administration window (see Figure 7.2). If you wish to view the Settings without making any changes, you may select **Read Only**. No password is required in read only mode.

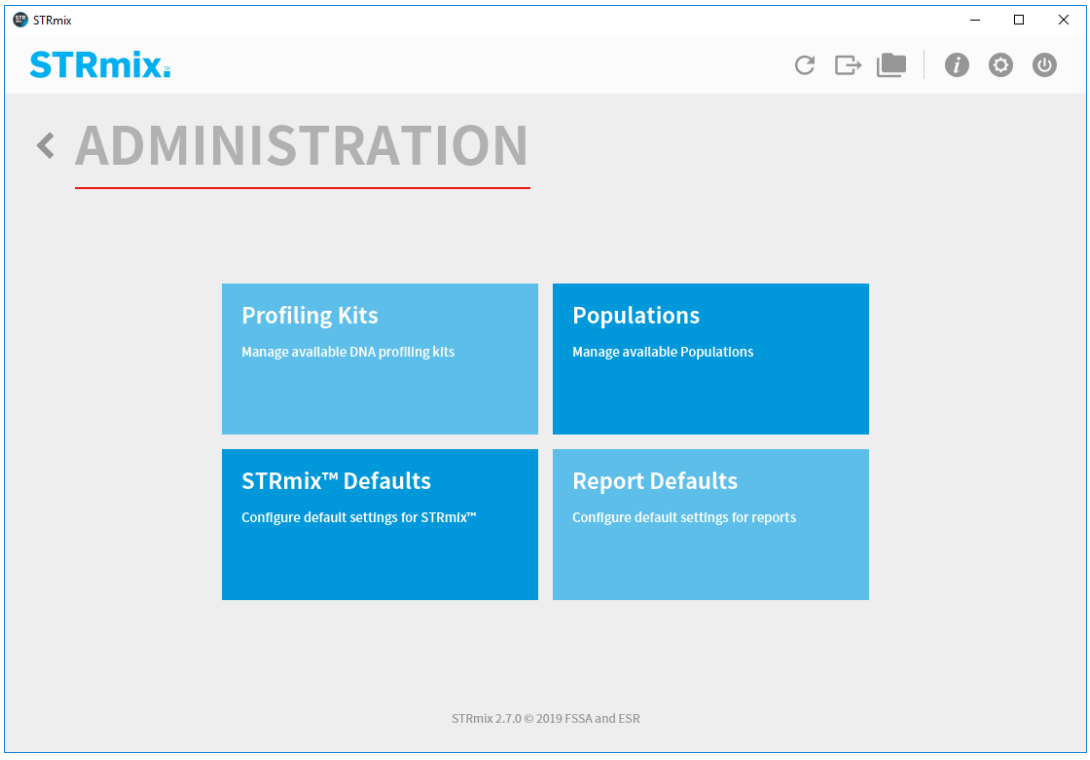


Figure 7.2 Main Settings window

From the Administration window you can add or edit profiling kits and populations, change the STRmix™ Default settings, and change the Report default settings. STRmix™ uses the information from within the kit and population files and STRmix™ default parameters file when interpreting DNA profiles and calculating *LR*s.

STRmix™ is prepopulated with some exemplar kits, population files and default parameter values. These are provided for demonstration and instruction purposes only. It is the responsibility of the individual laboratory to create and verify new files and parameters specific to their DNA profiling data prior to beginning DNA profile interpretation with STRmix™.

7.1 Default settings

Defaults for STRmix™ parameters used for the interpretation of single source and mixed DNA profiles can be set under **STRmix™ Defaults** (see Figure 7.3).

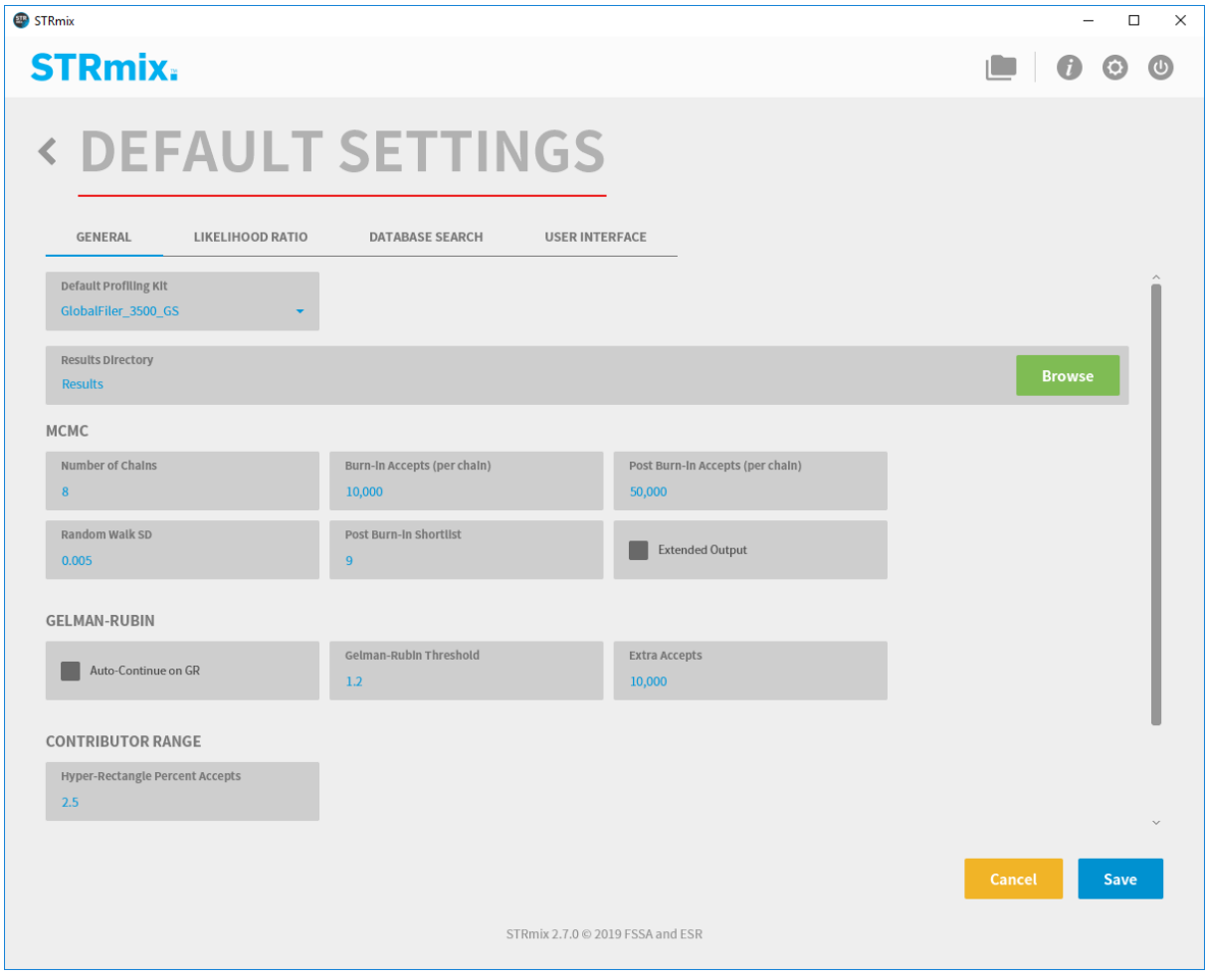


Figure 7.3 Default settings window

The STRmix™ parameters found both here on the Default Settings screen, and elsewhere in the software, are summarised in the table below within section 7.2. Many parameters are expected to be affected by laboratory specific variables e.g. STR kit, PCR and CE injection protocols, and instrumentation platforms. On occasion they may need to be adjusted for a specific sample. Some parameters are predetermined values defined during the developmental validation of STRmix™.

7.2 An explanation of the STRmix™ settings

General Settings	
Setting	**Explanation**
Default profiling kit	The STR profiling kit that is selected by default within a profile Interpretation.
Results directory	Folder location where all Results will be saved automatically. This defaults to the Results folder within the STRmix™ root directory.
Number of chains	STRmix™ default = 8. The number of MCMC chains to be used. Eight chains is recommended to effectively explore the probability space.
Burn-in accepts (per chain)	Default = 10,000. The number of burn-in accepts required per chain. If there are 8 chains run, there will be 80,000 Burn-in

	accepts. This variable can be changed for a specific run, without affecting the default.
Post burn-in accepts (per chain)	Default = 50,000. The number of post burn-in acceptances required per chain. An 8 chain run will result in 400,000 post burn-in accepts. This variable can be changed for a specific run, without affecting the default.
Random walk SD	Default = 0.005. RWSD sets the step size distributions for the random Gaussian walks.
Post burn-in shortlist	Default = 9.0. The $\log_{10}$(likelihood) that will be used to remove genotype sets from the post-burn-in list. All genotype sets that have a $\log_{10}$(likelihood) less than the set orders of magnitude of the current genotype set (and tested using the current values of mass parameters) are removed.
Extended Output	Default = off. Checking the box will turn Extended Output on. This will save all the accepted MCMC iteration information (and how many intervening rejections there were) into a separate file. Extended output is used for diagnostic purposes only and requires a large amount of data storage.
Auto-continue on GR	Selecting this allows the user to set a Gelman-Rubin value threshold. If this threshold is not met at the end of interpretation, further post burn-in accepts are triggered to allow for better convergence.
Gelman-Rubin Threshold	The Gelman-Rubin threshold value at which additional accepts are triggered if it is not met at the end of the initial analysis.
Extra Accepts	The number of extra post burn-in accepts run per chain if the Gelman-Rubin auto-continue is triggered. The analysis terminates after these extra iterations, regardless of the GR value at that point.
Hyper-rectangle percent accepts	The percentage of accepted post burn-in MCMC iterations to be included in hyper-rectangle within a varNOC interpretation. Default = 2.5%
Number of threads	For best performance the number of threads should be directly related to the number of cores or virtual cores available within your analysis computer. STRmix™ will automatically detect this on installation, however the user can set these manually as well.
Low Memory Mode	Option to conserve memory at the cost of extra execution time.
Likelihood Ratio Settings	
Default populations	Use the "**+**" button to add and the "**–**" button to remove default populations for LR calculations
Assign Sub-source *LR*	Select to turn on the sub-source *LR* calculation. Previously called the Factor of N!, this is also referred to as the two stain or two trace problem in the literature and helps account for contributor order within mixed DNA profiles if you are addressing a sub-source (or 'the POI is a contributor to the DNA in the mixture') level proposition. Deselecting this option will enable a sub-sub-source *LR* calculation (when for example you would be addressing a

	specific proposition such as 'the POI is the major contributor to a mixture of DNA').
Priors Method	Select Stratify to use a stratified *LR* within a varNOC interpretation. Select MLE to calculate an MLE *LR*. Default = Stratify
Calculate HPD	Select this option to enable the calculation of the HPD *LR* by default. Deselecting turns off this option. Default = on
MCMC uncertainty	When selected, STRmix™ will consider genotype set weights as distributions and resample from these distributions during the HPD calculation. Weights are modelled as having a gamma(EC, 1) distribution where EC is an effective count. The effective count is the number of independent iterations where the genotype set has been the focus of the MCMC and is based on the effective sample size calculation.
Allele frequency uncertainty	When selected, STRmix™ will consider allele frequencies as distributions and resample from the distributions during the HPD calculation.
Number of HPD iterations	Default = 1000. The number of iterations used within the Highest Posterior Density calculation to create the probability interval.
Probability interval quantile	Default = 99. The quantile used within the HPD calculation for the probability interval.
Probability interval sides	Default = 1. The number of sides used within the HPD calculation for the probability interval (1 or 2).
Database Search Settings	
Database File	**Browse** or drag and drop to select a default Database File to use for database searches.
Minimum *LR*	Default = 1,000,000. Set the list management value for Database Searches. Only profiles with *LR*s above the value entered here will be returned.
Population for Search	Choose from the drop-down to set the default population to use for Database Search.
Type of Search	Select Standard to carry out a standard Database Search. Select Familial to carry out a Familial Search.
F_{ST} for Search (Standard Search)	Default = 0.0b(1.0,1.0). Set the F_{ST} value to use for a standard Database Search.
Assign Sub-Source *LR* (Standard Search)	Default = on. Deselect to turn off the sub-source *LR* calculation in a standard Database Search.
Mutation Rate (Familial Search)	Default = 0.001. Set the mutation rate to use for a Familial Search.
Priors Method	Select Stratify to use a stratified *LR* within a varNOC interpretation. Select MLE to calculate an MLE *LR*. Default = Stratify

Auto Database Search	Default = off. Select to automatically carry out a Database Search at the end of each Interpretation. If Auto Database Search is selected, a Database File and Population for Search must be set.
User Interface Settings	
Default text file directory	Path to the folder where STRmix™ input files are located. When adding text files for interpretation this is the folder that STRmix™ will automatically open. Default = STRmix™ Results directory
Allow text sample renaming	Default = off. Check to allow the renaming of samples imported from text files.

7.3 Profiling kits

Refer to the STRmix™ Implementation and Validation Guide for further details on how to determine the parameter settings present in STRmix™ kit files.

Select **Profiling kits** from the Administration window to open the Manage kits window (see Figure 7.4 for a GlobalFiler™ example).

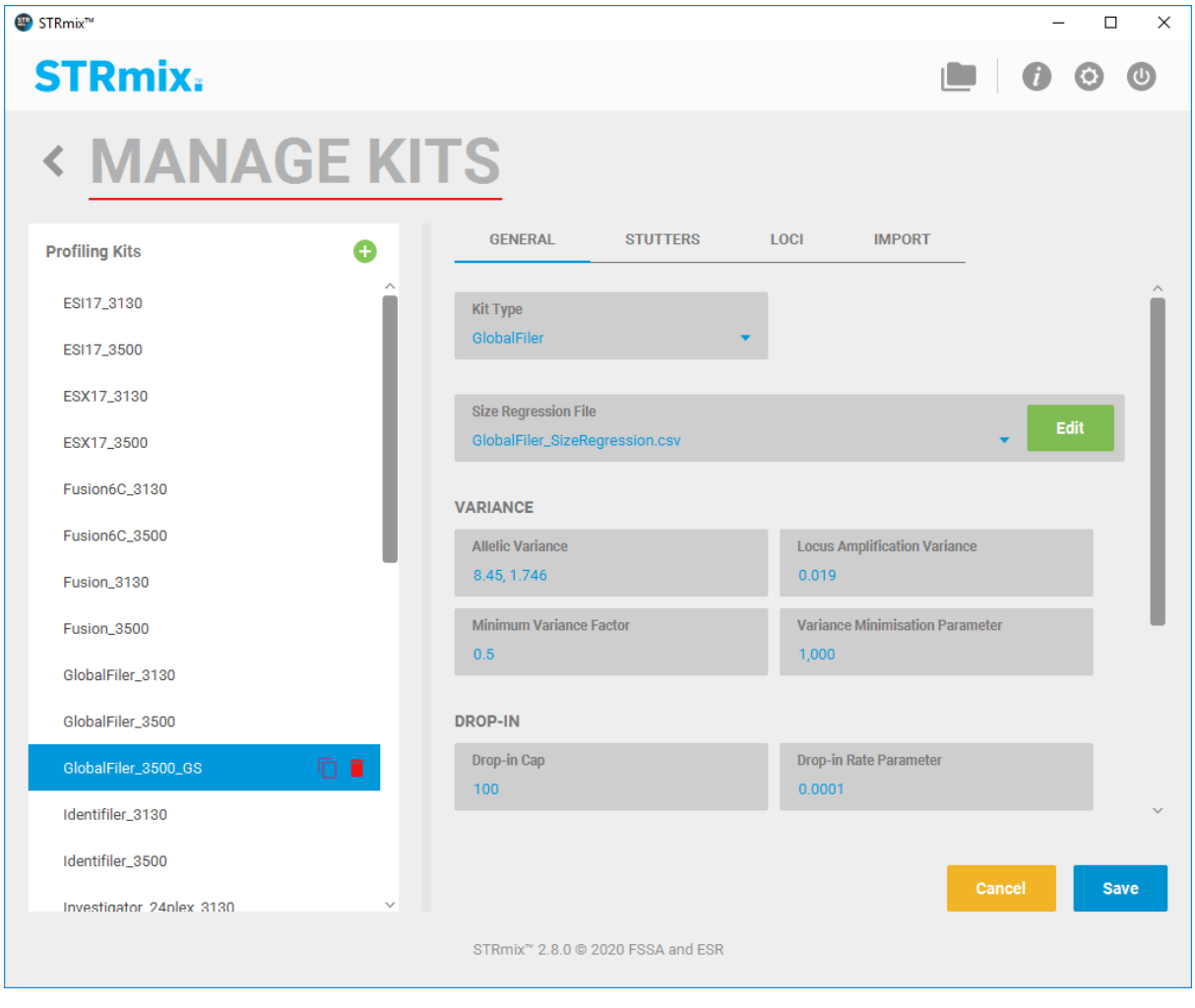
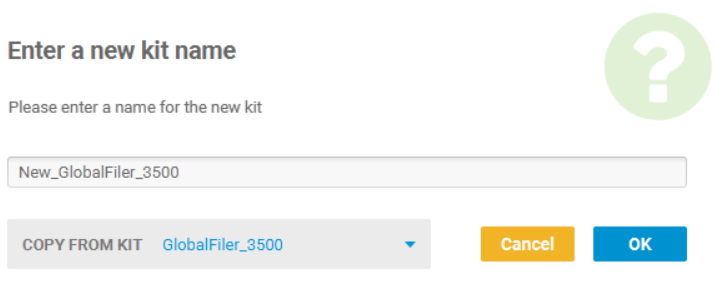


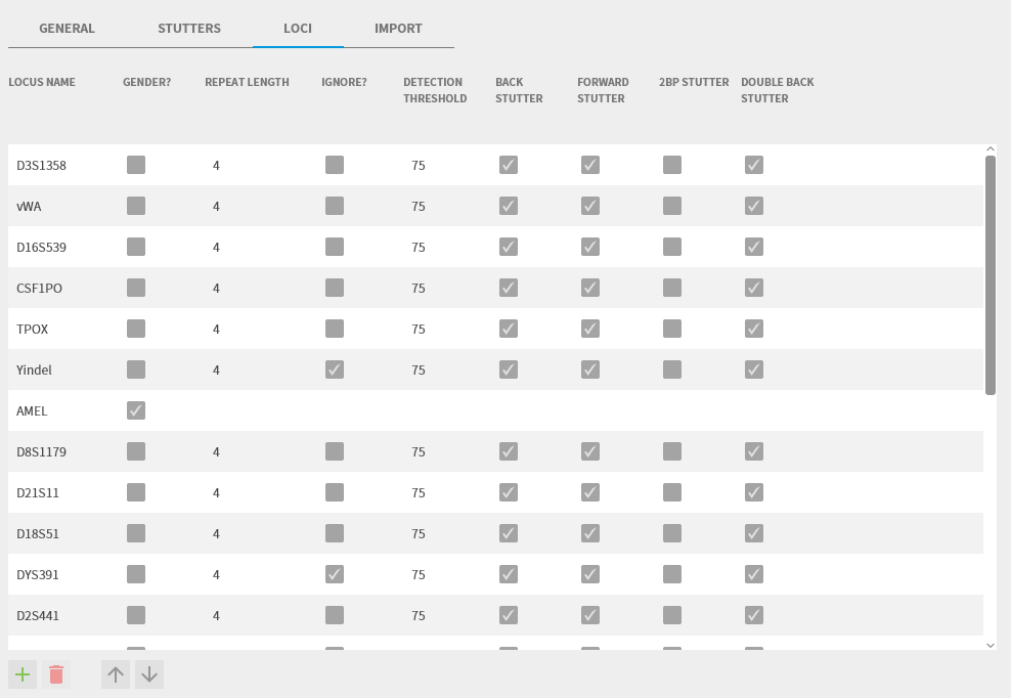
Figure 7.4 Manage DNA Profiling Kit window

7.3.1 Adding a new STR profiling kit

Select **Profiling Kits** from the Administration window (see Figure 7.2).

Step	Action
1	Select the "**+**" button to add a new kit (see Figure 7.4)
2	Enter a unique identifying name in the kit name field. If appropriate select an existing kit to copy information from (for example locus names) else, select **Don't copy from another kit** from the dropdown. 
3	Within the General tab, enter the kit settings as appropriate. Refer to Table 7-1 for a description of each setting.
4	Within the Stutters tab, enter the stutter kit settings as appropriate. Refer to Table 7-1 for a description of each kit setting. Select **Add New Stutter** to add a new stutter type. Select the pencil icon to edit the stutter type name. Select the rubbish bin icon to delete the stutter type. Select the arrows to move the stutter type position within the results. 
5	Within the Loci tab, set the per locus Detection threshold. Select each stutter type to be modelled for each locus, select **Gender** for gender loci, and **Ignore** to ignore a locus within a deconvolution. The Ignore Loci function can be used to ignore the non-autosomal STRs in the next generation multiplexes,

for example the Y-Indel and DYS391 marker within Applied Biosystems' GlobalFiler™.



6	To add loci, select "**+**" within the Loci tab. To delete loci, select the rubbish bin icon. Select the up and down arrows to change the order of loci in the kit.
7	Select **Save** to save or **Cancel** to return to the Administration window without saving.

7.3.2 Importing optimised variance values from a Model Maker interpretation

Once a Model Maker interpretation has been completed, to prevent transcription errors, the optimised variance parameters can be imported into the kit using the Import tab within Profiling kits. The parameters that will be imported are: Allelic Variance, Locus Amplification Variance (rounded to 3 decimal places), and all relevant Stutter Variance values.

Within the Profiling kit, select the Import tab (see Figure 7.6) and drag and drop or **Browse** to find the Model Maker results. Select **Import**. The user will be prompted to confirm that they wish to overwrite the kit settings with the Model Maker results. Select **No** to cancel or **Yes** to confirm (see Figure 7.5). Once kit settings are updated, the kit can be saved or the changes cancelled.

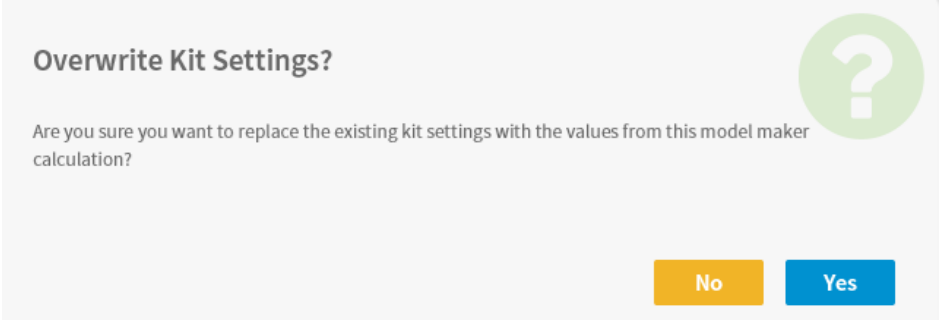


Figure 7.5 Confirmation message to overwrite kit settings.

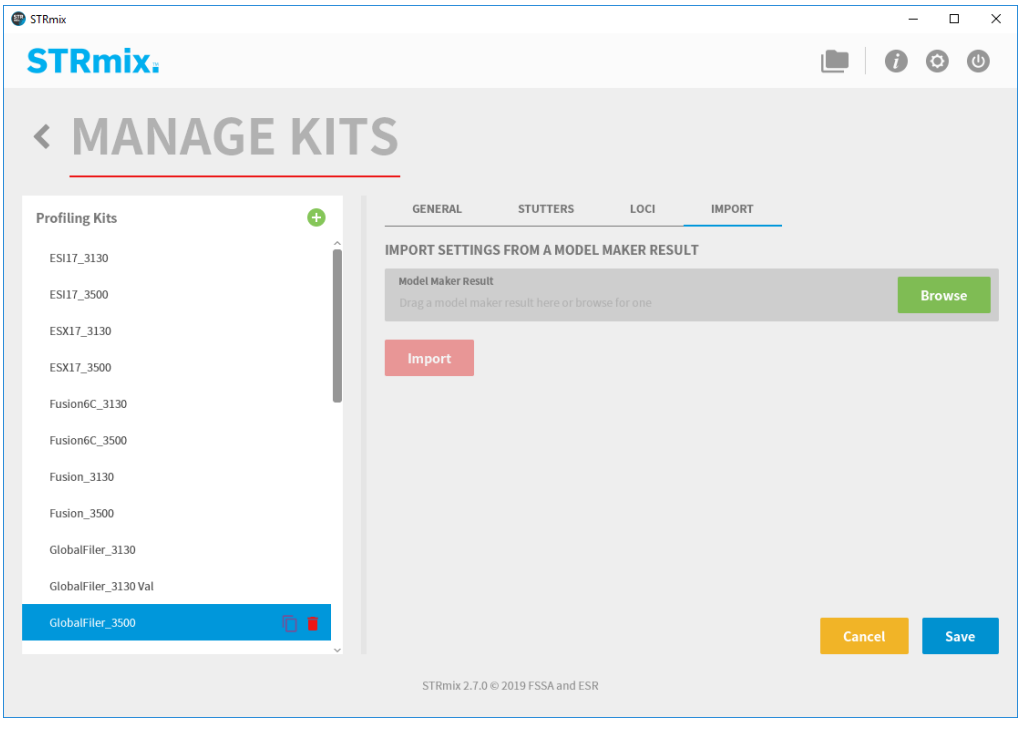


Figure 7.6 Import tab within Profiling kits

Notes

The kit being updated with the Model Maker optimised variance values must have the same name as the kit that was used in the Model Maker interpretation.

The kit being updated must have the same stutter types as the kit that was used in the Model Maker interpretation.

Table 7-1 Kit settings

General	
Kit type	Select the type of kit. Common kits are pre-entered. If the kit type is unknown, select **Unknown** from the drop down list. The kit type is used to plot the LSAE settings and is used within the CODIS report. The

	CODIS report will not generate if a kit type is not selected and the LSAE plot will be in black and white, not colour, if no kit type is selected.
Size regression file	The size regression file is used to calculate the size in bp of alleles not present in an input file (for example if conditioning on an individual who has alleles missing or 'dropped out' of the evidence input file).
Allelic Variance	Parameters that describe how variable allele peak heights are. Allelic variance (c^2) is modelled using a $\Gamma(\alpha_1, \beta_1)$ prior. α_1, β_1 are determined by the laboratory by use of Model Maker.
Locus amplification variance	A variable that penalises loci which have locus amplification efficiencies that differ from the mean of the other loci values. The Locus Amplification Variance parameter is determined using Model Maker and is expected to be affected by laboratory specific variables.
Minimum variance factor	Default = 0.5. The minimum allowable value the allele and stutter variance constants can take in relation to the mode of their prior distributions. For example, if the mode of c^2 = 4.2 and the minimum variance parameter setting = 0.5, then the smallest value the variance constant can take is 4.2 x 0.5 = 2.1.
Variance minimization parameter	Default = 1000. Used within the calculation of the log(likelihood). Refer to the discussion of the Taylor Quantum Effect from the STRmix™ User's Manual. If *b*=0, STRmix™ returns the original model (from within V2.5 and prior).
Drop-in parameters	Four parameters are involved in the drop-in modelling in STRmix™ V2.6. These are: 1. *Z*: the analytical threshold 2. *Y*: a cap on the maximum allowed drop-in height per allele (drop-in cap) 3. Drop-in parameter Pr(C) where C signifies the event of drop-in whether above or below the threshold. Pr(C) will therefore be larger than the observed drop-in rate 4. Drop-in distribution parameters, α_3, β_3. Two parameters of a gamma function. Selecting Uniform will apply a uniform drop-in frequency penalty.
Maximum degradation	Default = 0.01. The maximum allowable degradation for any one contributor during the entire MCMC process. For a heavily degraded profile, the maximum degradation may be increased to say; 0.1.
Degradation start point	The point in bp where degradation is first applied to the profile. This should be set to the approximate start of your STR multiplex. Selecting use smallest peak means degradation will start being applied to peak heights at the smallest molecular weight peak observed in the evidence input file(s).
Saturation threshold	Observed peak heights within an electropherogram may be saturated if they are above the saturation threshold calculated for a CE instrument. This means that the peak's height is not accurately captured and therefore the observed stutter peak height for the observed saturated allele could be smaller than its expected value. For this reason,

	STRmix™ implements an alternative model for expected stutter when the corresponding expected allelic peaks are saturated. The value for Saturation is determined empirically and is expected to be specific to the model of electrophoresis instrument used.
Stutters (settings per stutter type)	
Stutter enabled	Stutter type is enabled
Position relative to parent	The position of the stutter relative to the parent in the form i, j where *i* is whole repeat units and *j* is base pairs. For example, -1,0 is one whole repeat unit less than the parent allele, 0,-2 is two base pairs less than the allele, and 1,0 is one repeat unit greater than the allele.
Inversely proportional to	Select to specify that the peak height variability is inversely proportional to parent allele height. Deselect to specify that the peak height variability model is inversely proportional to the expected stutter peak height.
Maximum stutter ratio	The maximum allowable back stutter proportion permitted, i.e. 0.30 = 30%. Selecting No maximum turns this filter off. The maximum stutter ratio parameter should be determined by the laboratory by empirical trials and is expected to be affected by laboratory specific variables.
Stutter Variance	Parameters that describe how variable stutter peak heights are. Stutter variance (k^2) is modelled using a $\Gamma(\alpha_2, \beta_2)$ prior. α_2, β_2 are determined by the laboratory by use of Model Maker.
Stutter regression file	Link to the stutter regression file for calculating expected stutter height.
Stutter exceptions file	Link to the stutter exceptions file containing stutter ratios used for calculating expected stutter height.
Loci	
Gender	Used to select loci within the kit to designate as Gender loci (eg Amelogenin). Gender loci will not be included in calculations.
Repeat Length	Repeat length of the locus. Used for stutter modelling.
Ignore	Used to ignore loci within the kit. The Ignore Loci function can be used to ignore the non-autosomal STRs in the next generation multiplexes, for example the Y-Indel and DYS391 marker within Applied Biosystems' GlobalFiler™
Detection Threshold	The per locus minimum analytical threshold, *Z*, (in rfu) used to read the epg. The Detection Threshold should be determined by empirical trials and is expected to be affected by laboratory specific variables including data analysis software. *Z* is locus dependent. The analytical threshold set in STRmix™ should align with the analytical threshold used in epg analysis by the lab. Select **Detection Thresholds** from the **Manage DNA Profiling kit** function to open a Set Detection Thresholds window. The STRmix™ detection threshold setting does not limit the input data to peak heights above this value, however pre-checks will warn the user if there are peaks present below this value (see 3.3.9.1). It is used within the calculation of the dropout probability.

7.3.3 Editing an existing STR profiling kit

Select **Profiling Kits** from the Administration window (see Figure 7.2).

Step	Action
1	Select the kit from the Profiling kits dropdown list (see Figure 7.4)
2	Edit the relevant details as appropriate. Select **Edit File** in the Stutter Regression File or Stutter Exceptions File to open the files for editing (refer below for more details)
3	Select **Save** to save or **Cancel** to return to the Administration window without saving.

7.3.4 Deleting an STR profiling kit

Select **Profiling Kits** from the Administration window (see Figure 7.2).

Step	Action
1	Ensure the appropriate kit is selected in the Profiling Kits window
2	Select **Delete Kit** (rubbish bin icon)
3	Select **Delete** to confirm or **Cancel** to cancel.

7.3.5 Creating and editing Stutter files

Stutter files can be edited within STRmix™ or outside STRmix™, either in Notepad or in MS Excel and then saved in the STRmix™ directory Stutter folder in the appropriate file format. If editing a Stutter file in Excel, you will need to convert the text to columns and save as .csv file. If stutter files are edited outside of STRmix™ you will need to enter the Administrator password to have these changes accepted after entering and exiting the Main Settings window (if your settings are protected via a password).

The format of the stutter and forward stutter files is; locus name, Y-intercept, and slope. The locus names must be identical to those within the kit. The intercept and slope parameters for each locus are used to create an expected stutter ratio for each allele at that locus. In Figure 7.7, an example of a stutter file for Fusion™ is provided.

```
Locus,Intercept,Slope
D3S1358,-0.044,0.00795
D1S1656,0.0185,0.00421
D2S441,0.0615,-0.00111
D10S1248,-0.0595,0.00966
D13S317,-0.0488,0.00869
Penta E,-0.0172,0.00395
D16S539,-0.0581,0.0108
D18S51,-0.0393,0.00765
D2S1338,-0.0139,0.00449
CSF1PO,-0.0493,0.00956
Penta D,-0.0241,0.00377
TH01,0.000254,0.00322
vWA,-0.0881,0.00964
D21S11,-0.0366,0.00372
D7S820,-0.0672,0.0115
D5S818,-0.0436,0.0087
TPOX,-0.0292,0.00582
DYS391,0,0
D8S1179,0.0131,0.00396
D12S391,-0.0878,0.00904
D19S433,-0.075,0.01
FGA,-0.0222,0.00383
D22S1045,-0.117,0.0132
```

Figure 7.7 Example of a Fusion™ Stutter file

An example of part of a stutter exceptions file is provided in Figure 7.8. The stutter exceptions are the stutter ratios based on values determined for the longest uninterrupted sequence (*LUS*) of core repeats within an allele. Stutter ratio has been shown to be linearly related to *LUS* [7, 8]. The exceptions file contains the 'per allele' stutter ratios calculated for *LUS* values but presented in the file against allele designation. The stutter exceptions file also contains the average stutter ratios per allele for loci where an average of the stutter ratios at each allele was found to best explain the observed stutter ratios for the locus. Refer to the STRmix™ Implementation and Validation Guide for further details on how to create stutter regression and exceptions files.

	A	B	C	D	E	F
1	Allele	D3S1358	vWA	D16S539	CSF1PO	TPOX
2	2.2	0	0	0	0	0
3	3	0	0	0	0	0
4	4	0	0	0	0	0
5	5	0	0	0	0	0
6	6	0	0	0	0	0
7	6.3	0	0	0	0	0
8	7	0	0.02473	0	0	0
9	7.3	0	0	0	0	0
10	8	0	0	0	0	0
11	8.3	0	0	0	0	0
12	9	0	0	0	0	0
13	9.3	0	0	0	0	0
14	10	0	0	0	0	0
15	10.1	0	0	0	0	0
16	10.3	0	0	0	0	0
17	11	0.02092	0.02473	0	0	0
18	11.3	0	0	0	0	0
19	12	0.03078	0	0	0	0
20	12.1	0	0	0	0	0
21	12.2	0	0	0	0	0
22	13	0.04064	0	0	0	0

Figure 7.8 Example of part of a stutter exceptions file for GlobalFiler™

7.3.6 Creating new stutter regression and exceptions files

Step	Action
1	Open MS Excel. Following the example in Figure 7.7 (regression file) or Figure 7.8 (exception files) enter the header row
2	Stutter Regression Files: On each new row enter the full Locus name (identical to kit locus name) and Intercept and Slope values for that locus (utilising columns A, B and C respectively) Exceptions Files: In column A on each new row enter the allele value. Enter the stutter ratio (see details for how to calculate this in the STRmix™ Implementation and Validation Guide) for each known allele value for each locus. Enter 0 values for missing, unknown or negative values.
3	Save the appropriately named file as .csv in the Stutters folder within the STRmix™ directory
4	If you are using password protection, for this change to take effect, you will need to enter the Administration password and confirm the change or, select the **Reload files on disk** icon  from within the Administration main window.

Notes

The values used for stutter regression and exceptions files should be taken from empirical data. If using a per-locus value (e.g. for forward stutter files), the average stutter ratio observed should be entered as the intercept value and the slope set to zero. If using a per-allele value, intercept and slope values are determined by regressing the stutter ratio (or forward stutter ratio) against allele. The intercept and slope are taken from the line of best fit. See the STRmix™ Implementation and Validation Guide for more details.

Stutter ratio values within exception files are estimated by regressing *SR* against LUS or taking the average of the observed values. Per allele *SR* values are then calculated and transcribed into the exceptions file. If a value is not available for a particular allele the *SR* is taken from the per allele rate from the stutter file.

In the absence of a calculated *SR* value greater than 0.001 from either the Stutter Exceptions or Stutter files, STRmix™ will return a maximum value of 0.001 for *SR*.

7.4 Populations

Select **Populations** from the Administration window to open the Manage Populations window (see Figure 7.9).

7.4.1 Adding a new population

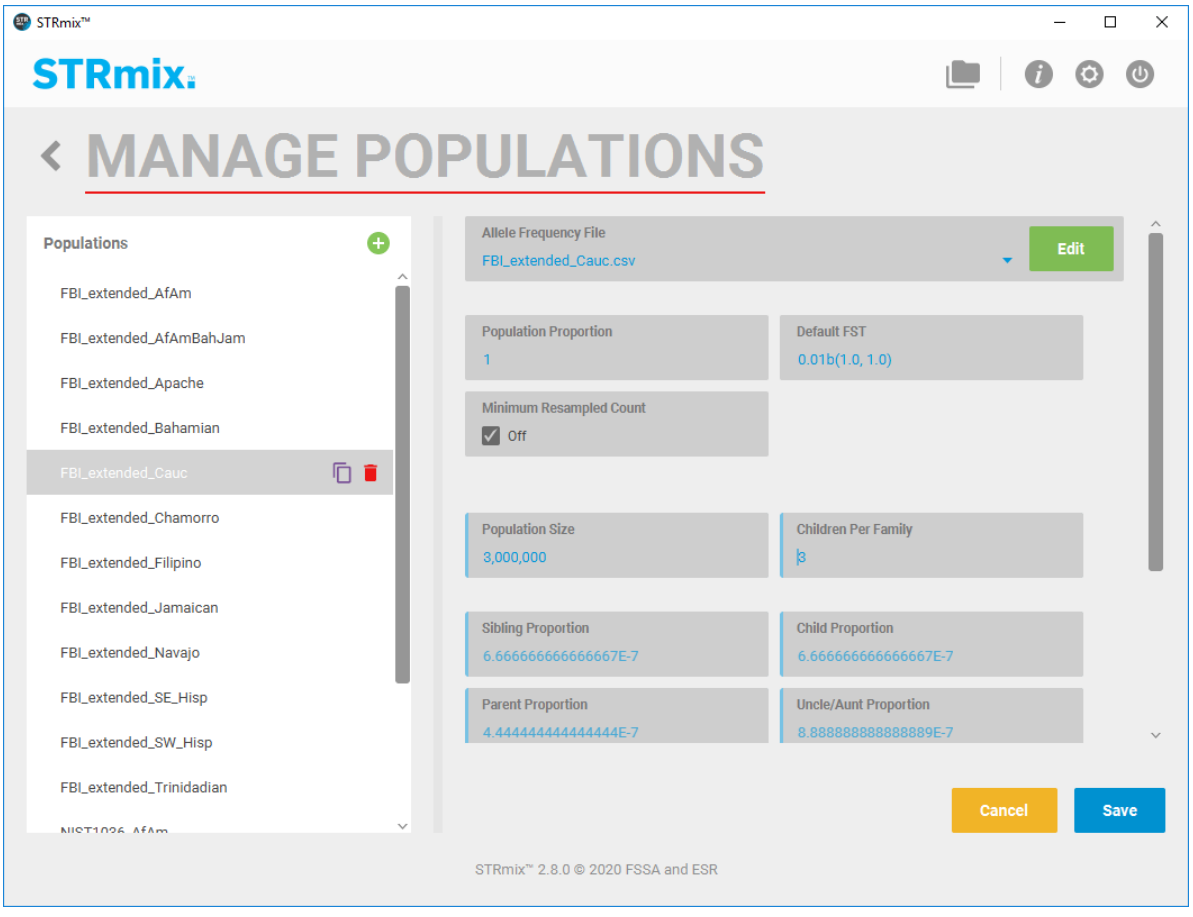


Figure 7.9 Adding a new population window

Step	Action
1	Select the "**+**" button to add a new population
2	Enter a unique identifying name in the population name field. If appropriate select an existing population to copy information from, otherwise select **Don't copy from another population** from the dropdown. 
3	Select an **Allele frequency file** in the Allele Frequency File field from the dropdown list. Select an appropriate allele frequency file (refer below for more details)
4	Enter an appropriate Population Proportion value used within the calculation of a stratified *LR*

5	Enter the Default F_{ST} value for the population. The uncertainty in the value for F_{ST} (or theta) is modelled using a $\beta(\alpha_4, \beta_4)$ distribution. To use a fixed estimate *x* of F_{ST} enter *x*b(1,1). As an example, 0.01b(1,1) will use an F_{ST} value of 1%.
6	Enter the Minimum Resampled Count. By default this is set to Off. If a custom value is entered it must be a number greater than zero.
7	If a unified *LR* is intended to be calculated, enter the population size and average number of children per family. These values can be estimated from a census for example. The remaining fields of the population (prior probabilities of each population) will auto populate
9	Select **Save** to save or **Cancel** to return to the Administration window without saving.

Notes

The Minimum Resampled Count (MRC) is designed to prevent large differences between the sub-source and HPD *LR*s due to alleles not previously observed within the population. If MRC = Off then there is no minimum allele count in the HPD. If MRC>0 then the minimum allele count is the maximum of the random gamma or the MRC. If wanting to use a Minimum Resampled Count, then we suggest using 1 as a value, however we recommend that this is validated prior to use as with all user settings. In V2.8, the population size is no longer additionally being used to set a minimum allele frequency (for the point estimate or HPD *LR*).

7.4.2 Editing an existing population

Step	Action
1	Select **Populations** from the Administration window to open the Manage populations window (see Figure 7.9)
2	Select the appropriate population from the Populations window
3	Edit the relevant details as appropriate. Select **Edit** in the Allele Frequency File field to open the Allele Frequency file for editing (refer below for more details)
4	Select **Save** to save or **Cancel** to return to the Administration window without saving.

7.4.3 Deleting a population

Step	Action
1	Select **Populations** from the Administration window to open the Manage Populations window.
2	Select the appropriate population from the Populations window
3	Select **Delete Population** (rubbish bin icon)
4	Select **Delete** to delete the population or **Cancel** to return to the Manage Populations window.

7.4.4 Creating and editing allele frequency files

Allele frequency files can be edited within STRmix™ or outside STRmix™, either within Notepad or Excel. The instructions provided below in section 7.4.5 Editing existing allele frequency files describe how to edit allele frequency files from within STRmix™. Allele frequency files are saved within the AlleleFreq folder within the STRmix™ root directory.

7.4.5 Editing existing allele frequency files

Step	Action
1	Select **Populations** from the Administration window to open the Manage populations window
2	Select the appropriate population and select **Edit** against the Allele Frequency File field to open the selected allele frequency file in Excel for editing (see Figure 7.10)
3	Edit the file as appropriate
4	Close the file and select **Save** when prompted to save the file or **Don't Save** to close without saving.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA
2	1	0	0.000121	0	0	0.000121	0	0.000424	0.000242	0.000121	0	0.000242
3	4	0	0	0	0	0	0	0	0	0	0	0
4	5	0	0	0	0	0	0.002546	0	0	0	0	0
5	6	0	0	0	0	0	0.223509	0	0	0	0	0
6	6.3	0	0	0.000364	0	0	0	0	0	0	0	0
7	7	0	0	0.018004	0.000364	0	0.185378	0.000242	6.06E-05	0	0	0
8	7.3	0	0	0	0.000121	0	0	0	0	0	0	0
9	8	0.018914	0	0.153067	0.002789	0	0.104207	0.119908	0.016368	0	0	0
10	8.1	0	0	6.06E-05	0	0	0	0	0	0	0	0
11	8.3	0	0	0	0	0	0.000606	0	0	0	0	0
12	9	0.012852	0	0.163797	0.02346	0	0.140034	0.075412	0.114998	0	0	0
13	9.1	0	0	6.06E-05	0	0	0	0	0	0	0	0
14	9.3	0	0	6.06E-05	0	0	0.334869	0	0	0	0	0
15	10	0.101419	0	0.263822	0.270187	0.000242	0.008729	0.065652	0.057772	0	0.000485	0
16	10.1	0	0	6.06E-05	0	0	0	0	0	0	0	0
17	10.3	0	0	0	0.000364	6.06E-05	0	0	0	0	0	0
18	11	0.076867	0	0.206656	0.297951	0.001516	0	0.299042	0.307165	0	0.003092	0
19	11.1	0	0	0	0	0	0	0	0	0	0	0
20	11.2	0	0	0	0	0	0	0	0	0	6.06E-05	0
21	11.3	0	0	0.000121	0	0	0	0	0	0	0	0
22	12	0.139185	0	0.15137	0.318986	0.000485	0	0.284675	0.29783	0	0.079595	0.000242
23	12.1	0	0	0	0	0	0	0	0	0	0.000546	0
24	12.2	0	0	0	0	0	0	0	0	0	0.00097	0
25	13	0.324988	0	0.035039	0.068016	0.003092	0	0.10742	0.178225	0.000424	0.238906	0.000667

Figure 7.10 Example of an Identifiler™ allele frequency file

7.4.6 Creating new allele frequency files

Allele frequency files can be created from the Excel tables commonly provided with modern population data publications. Alternatively, they can be manually created in Excel and saved as a .csv file. The instructions provided below are for creating allele frequency files in Excel. An example file is provided in Figure 7.11. Allele frequency files are saved within the AlleleFreq folder within the STRmix™ root directory. Note that allele frequencies should sum

to 1. If using an allele frequency file where the sum of the frequencies for any locus is greater or less than 1 by 0.05 (1±0.05) a warning will be given during the Interpretation setup.

Step	Action
1	Open a new Excel file. In column A, under the heading Allele enter all the observed alleles across all loci within the population. See Section 2 Allowable and rare alleles for a discussion about allowable and rare alleles.
3	In row A, add all of the locus names. These must be identical to the locus names within the kit but not necessarily in kit order.
3	For each locus enter the calculated allele frequency against the allele value. Enter 0 when no alleles were observed
4	In the last row enter N, the total number of alleles observed in the dataset for each locus
5	Save the appropriately named file in the AlleleFreq folder within the STRmix™ directory. If you are using password protection, for this change to take effect, you will need to enter the Administration password and confirm the change or, select the **Reload files on disk** icon from within the Administration main window.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Allele	CSF1PO	D10S1248	D12S391	D13S317	D16S539	D18S51	D19S433	D1S1656	D21S11	D22S1045	D2S1338	D2S441	D3S1358	D5S818	D6S1043	D7S820	D8S1179
2	2.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	3.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	4.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	6.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0028	0	0.0277	0
10	8	0.0055	0	0	0.1205	0.018	0	0	0	0	0	0	0	0	0.0055	0.0014	0.144	0.0139
11	8.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0014	0
12	9	0.0139	0	0	0.0776	0.1066	0	0	0	0	0	0	0.0014	0	0.0416	0	0.1676	0.0055
13	9.1	0	0	0	0	0	0	0	0	0	0	0	0.0014	0	0	0	0	0
14	9.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	10	0.2202	0	0	0.0471	0.0568	0.0083	0.0014	0.0028	0	0	0	0.2105	0	0.0554	0.0166	0.2562	0.1025
89	33.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
90	33.2	0	0	0	0	0	0	0	0	0.0263	0	0	0	0	0	0	0	0
91	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
92	34.2	0	0	0	0	0	0	0	0	0.0042	0	0	0	0	0	0	0	0
93	35	0	0	0	0	0	0	0	0	0.0014	0	0	0	0	0	0	0	0
94	36	0	0	0	0	0	0	0	0	0.0014	0	0	0	0	0	0	0	0
95	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
96	38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
97	39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
98	43.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
99	N	722	722	722	722	722	722	722	722	722	722	722	722	722	722	722	722	722

Figure 7.11 Example of an allele frequency file created in Excel

7.5 Report defaults

The default report settings are found within the Administration menu and then by selecting **Report Defaults** (see Figure 7.2). Within this all reports can be configured. Select the **Calculation Type** from the dropdown list and then select by highlighting the report in the **Reports to Run** window (see Figure 7.12). A list of configurable reports and report components available in STRmix™ V2.8 is given in Table 7-2. The order that each component appears in the report can be configured by dragging and dropping the different components into place. Different components can optionally be included in the report by checking the

Include? check box. Components can also be forced to appear on a new page within the report by checking the **New page?** check box.

The tick boxes next to the report name within the Reports to run window enables the report to be automatically generated at the completion of the analysis. If these options are disabled the user is able to generate reports retrospectively through the Reports window (see Section 6).

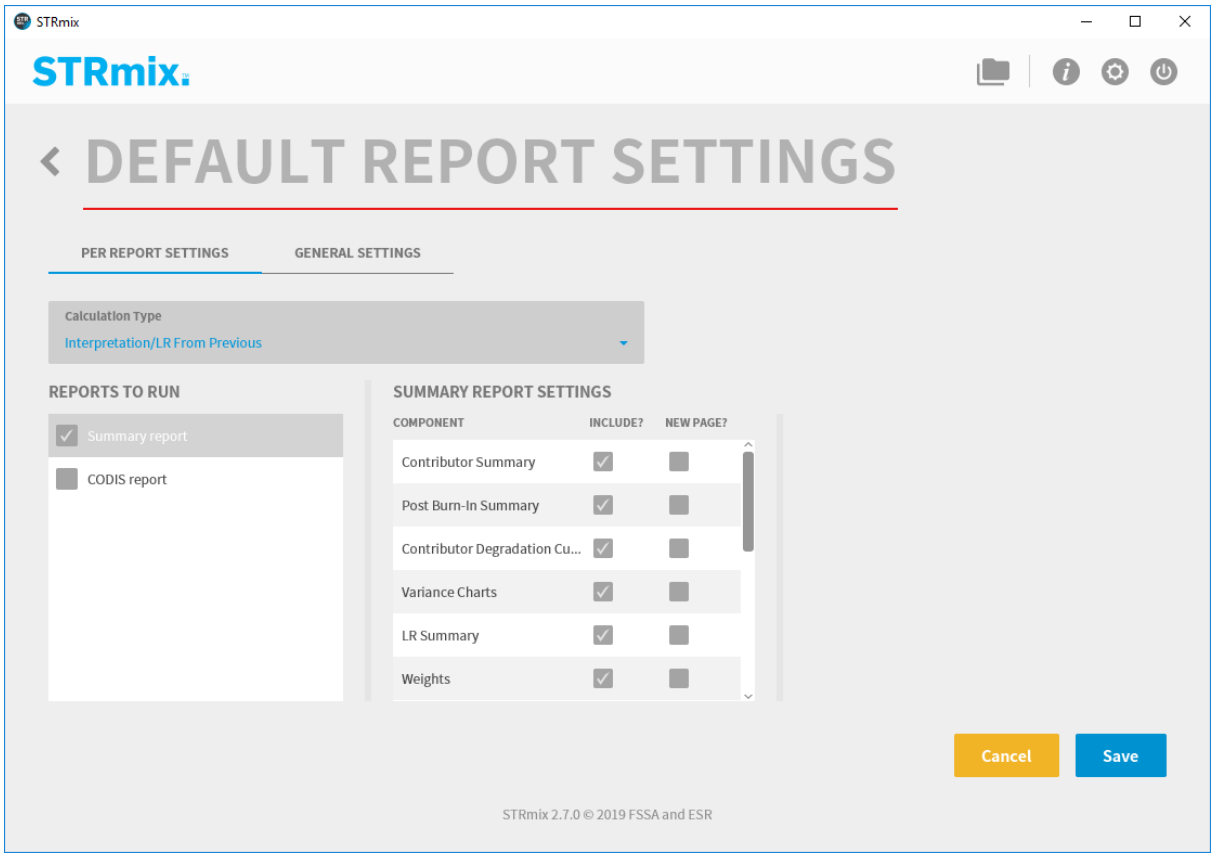


Figure 7.12 Configure report defaults screen

Table 7-2 Summary of configurable reports and report settings

Calculation type	Reports to run	Report settings
Interpretation/LR from previous	Summary report	Optionally include: Contributor Summary; contains template, mixture proportion, degradation and contributor order information. Post Burn-in Summary Contributor Degradation Curves Variance charts LR summary; displays the Summary of *LR* table and the Per Locus Likelihood Ratios table. Optionally display Unified *LR* and Population Stratified *LR* and *LR* summary

		for each N within a varNOC interpretation, and varNOC contributor *LR* summary Weights; set weight threshold for genotypes to be highlighted in the report (default = 0.99), set the limit for the number of printed genotypes per locus (default = -1, where all genotype combinations are printed) Component interpretation; set the label for components that fail to reach the genotype weight threshold, show the component interpretation for each individual contributor, group the component interpretation by contributor versus by locus Locus efficiencies Evidence files; displays the Locus, allele, size and height from peaks with the evidence input file(s) Evidence Peak Issues Reference files; displays the locus and alleles for the reference input file(s) Settings; records all run settings used in analysis Kit files; records all kit settings used in analysis Interpretation details; details for the previous interpretation if undertaking an *LR* from previous.
	CODIS report	Refer to Section 7.5.1for a discussion
Database search	Summary report	Sort *LR* by highest to lowest. Enable this option to sort the results from Database search by *LR* from highest to lowest.
Hd True Tester	Summary report	No configurable options
Model Maker	Summary report	No configurable options
Mix to Mix	Summary report	No configurable options
Top Down Approach	Summary report	Sort *LR* by highest to lowest. Enable this option to sort the results from Database search by *LR* from highest to lowest.

General report settings can be configured from the **General Settings** tab (see Figure 7.13). These settings are applied to all reports.

An optional laboratory logo can be applied to all pdf report types to the top left corner of the report. To add a logo, drag and drop an image (.png or .jpg) into the Lab Logo field or navigate to the image by selecting the **Browse** button in the Lab Logo field. To remove the logo, right-click in the **Lab logo** field and select **Clear**.

The report is formatted to fit the size of the paper selected, within the **Page Size** field, either A4 (210 by 297mm) or US Letter (215.9 by 279.4 mm).

The bottom margin can be adjusted by entered a number in the **Bottom Margin** field. The number represents pdf pixels. For example, a value of 50 brings the footer up around 3 cm and leaves a blank margin below it.

Select the **Date Format** from the dropdown list; either dd MMM yyyy or MMM dd yyyy.

Check **Auto-Open report** to open the report upon completion of the STRmix™ analysis. This setting applies to all pdf report types.

Check **Include Audit Info** to add audit hash information to the pdf reports. This setting applies to all pdf report types.

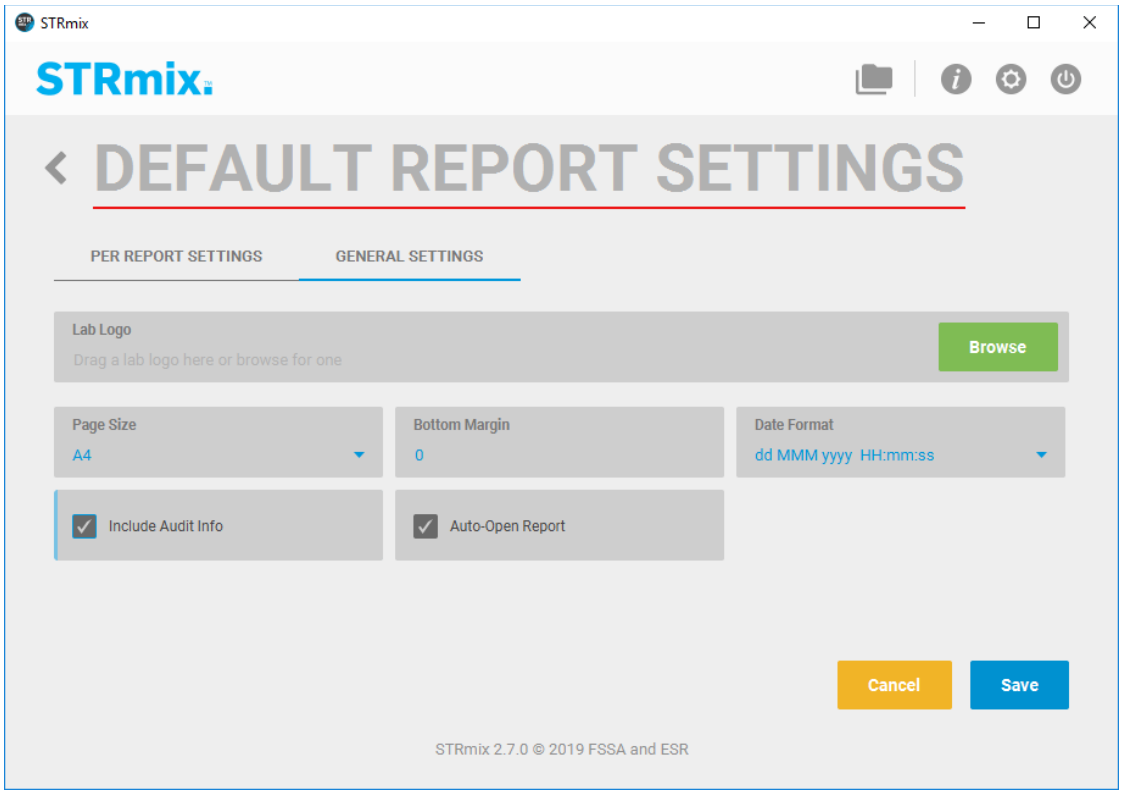


Figure 7.13 General settings tab within Default Report Settings

7.5.1 CODIS report

The CODIS report settings are used to configure the CMF report for CODIS upload. Check the box next to CODIS report in the Reports to run window (see Figure 7.14). Before running a CODIS report the settings must be configured. See Table 7-3 for a discussion of the different settings and logic behind the report. Note that if generating a CODIS report for a multiple kit interpretation, the CODIS report will only generate if all kits contain the same loci. The exception is that if all loci are contained within one kit (for example, all Identifiler™ loci are shared with the GlobalFiler™ multiplex), the loci within the larger kit will appear in the CODIS report.

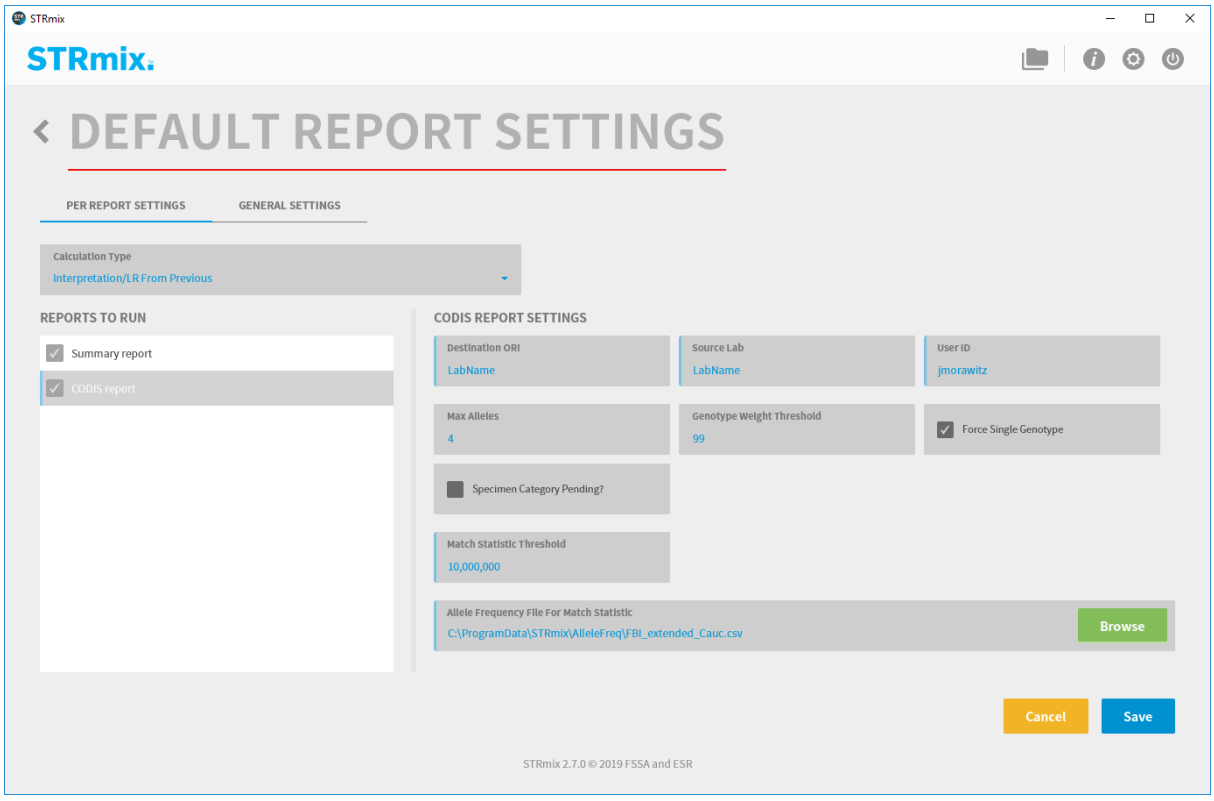


Figure 7.14 CODIS report settings window

Table 7-3 Description of the CODIS report settings

Destination ORI	CODIS ORI for the destination CODIS laboratory that will import the CMF file.
Source laboratory	The name of the source laboratory producing the CMF file. This could contain the name of the contract laboratory or be the same as the destination ORI.
User ID	Submit by User ID. If left blank, the username from the computer login will default into this field within Run Reports and can be edited.
Max Alleles	Defines the maximum number of alleles allowed per locus where the default is four and the maximum allowed is eight
Genotype Weight Threshold	Defines the weight threshold at which a genotype will be automatically included in the report. The default is 99%. The minimum genotype weight threshold is 51%. Decreasing the genotype weight threshold increases the number of eligible genotype combinations but lowers the stringency of the search
Force Single Genotype	Enabling force single genotype will mean that no other genotype combinations will be considered if one genotype combination is equal to or greater than the genotype weight threshold. If the setting is disabled, other alleles will be considered in the report up to the maximum number of alleles permitted.

Specimen Category Pending?	Default = off. Setting this will make the Specimen category default to Pending rather than Forensic, Unknown.
Match statistic threshold	Set the match statistic threshold above which a CODIS report will be generated for each contributor. Refer to the STRmix™ V2.8 User's Manual (Section 2.40) for a description of the calculations.
Allele frequency file for match statistic	Browse to the allele frequency file to be used for the calculation of the match statistic

Select **Save**. A CODIS xml report will be created for every contributor within the profile that meets the match statistic threshold. Each xml report will be labelled with the case name, sample name, "CODIS" and appended with the contributor number. CODIS xml files can be edited manually in a text editor such as Notepad or WordPad. In addition, a CODIS summary pdf report will be generated describing the interpretation details, CODIS CMF report settings, summary of contributors including eligible alleles, and match statistic.

8. MODEL MAKER

Model Maker is used to determine the variance constants for the different types of generalised stutter being modelled, allele peak height variance, and locus amplification variance; A^l, for different STR profiling kits, protocols and CE instruments of the same model (e.g. 3130). It is recommended that Model Maker be run for each kit and CE instrument model combination used within the laboratory. Note that instruments of the same model, e.g. multiple thermocyclers or CE machines do not need separate model maker analyses as they have been found to have minimal effects on peak height variability.

It is appropriate that these variances should be determined in a way that aligns with the way they will be used. For example, the variance in peak height supposes some distribution about an expected value. However, in empirical work this expected value is not available. It may be estimated from, say, dilution values, or a quantification value, however to make it fit for purpose it would be better if it were developed from the same mass values that will be used in the software. Accordingly, we adapt the MCMC approach to develop these parameters. A discussion on how Model Maker works is given in the *STRmix™ V2.8 User's Manual*.

8.1 Running Model Maker

NOTE: Model Maker is computationally intensive and may take many hours to complete analysis, especially as more stutter types are being modelled.

Select **Model Maker** on the Main Screen (refer Figure 3.2) to open the Model Maker module (see Figure 8.1). To undertake a Model Maker analysis two text input files are required. The first input file (the Single Source file) will contain single source evidence samples, with values retained for modelled stutter(s). The second input file (the Reference file) will contain the corresponding known reference profiles, with stutter removed at analysis. The sample names within the two files must be identical.

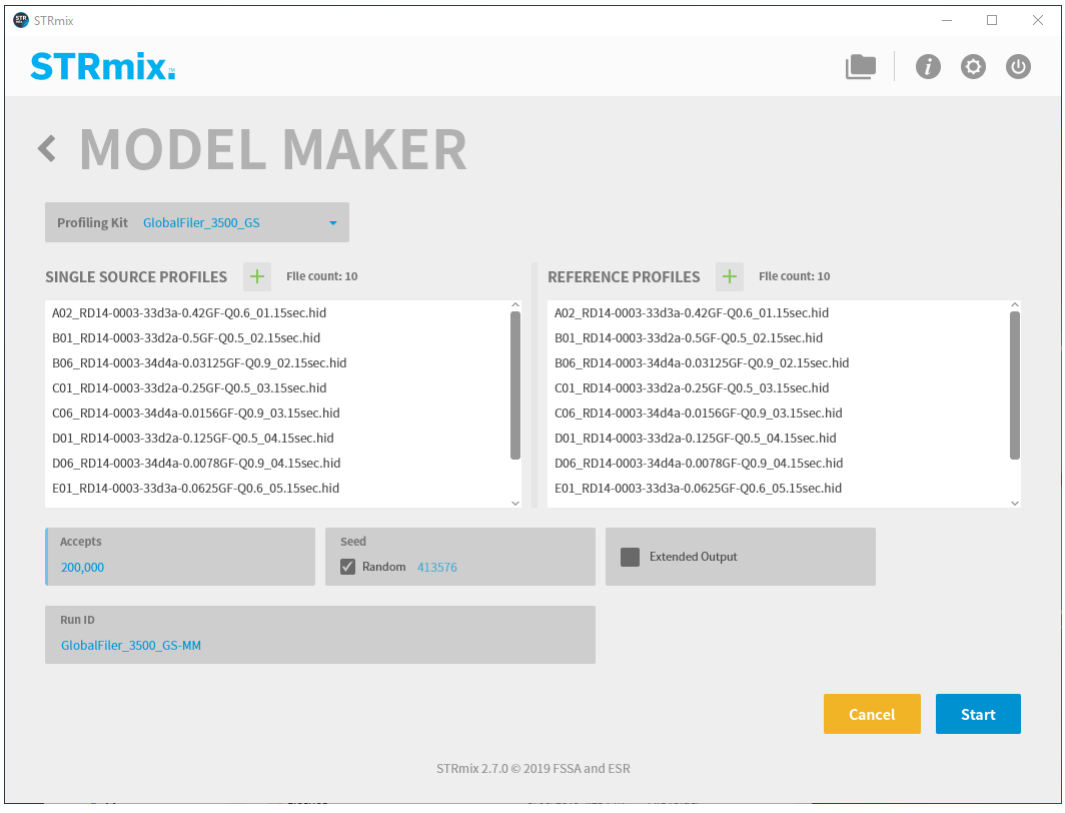


Figure 8.1 Model Maker window

Model Maker should be run with a range of samples of varying profile quality and peak heights. It is recommended that at least a full plate of data (approximately 90 samples) is analysed for each protocol/kit/instrument combination. This set is termed the training set as it is used to train the parameters. This analysis is only required to be undertaken once prior to using STRmix™ and after any significant material change to protocol or instrumentation which effects peak heights.

Step	Action
1	Select the appropriate kit file from the **Profiling Kit** drop down menu. Unlike previous STRmix™ versions, the number of allele columns do not have to be identical between the reference and the single source data.
2	Drag and drop the input file containing the crime samples into the **Single Source Profiles** box. Alternatively, select the **Add** icon "**+**", located next to **Single Source Profiles** to open the file selection window and navigate to the text input file containing single source DNA profiles containing Allele, Height, and Size for alleles and all types of stutters.
3	Drag and drop the input file containing the reference samples into the **Reference Profiles** box. Alternatively, select the **Add** icon "**+**", located next to **Reference Profiles** to open a file selection window. Navigate to the text input file containing single source reference DNA profiles containing Allele only.
4	When the files have been selected and successfully validated the list of sample names will be displayed in the Model Maker windows (see Figure 8.1).
5	Select **Start** to start Model Maker or **Cancel** to return to the main window.
6	If peaks below the analytical threshold are detected in the Single Source Profiles input file, a warning message will appear (See Section 3.3.9.1) Select

7	On completion, open the results by selecting the Open Model Maker Directory icon:  in the upper right third of the Summary window

Notes

There is no limit to the number of samples that can be in the text input files. The only limitation is that they must contain at least eleven STR peaks (this excludes Amelogenin). All reference profiles must be fully represented at all loci regardless of how partial their corresponding single source file profiles are.

Model Maker uses the default values that have been defined for detection threshold, stutter (input files and maximum allowed values) drop-in and saturation. Prior to running Model Maker, the default values for each of the utilised parameters should be updated to reflect the DNA profiling kit and Genetic Analyser from which the data to be analysed was generated. If Model Maker is to be run using a new profiling kit, the details that are applicable to this kit (including the stutter and stutter exception files) need to be entered into STRmix™ prior to starting the run (refer section 7.3 Profiling kits).

The default number of Accepts has been set to 200,000.

Model Maker optimises the stutters, allele and LSAE variance for each profile individually. Any profiles with less than 11 datapoints or with peaks above the saturation level (from the kit settings) are not included in the Model Maker analysis. Once optimised, each set of variance constant values (allele variance and stutter variances) have a gamma curve $\Gamma(\alpha,\beta)$ fitted with the criteria $\alpha > 1.5$ (to prevent exponential curves being fit). The LSAE variance is fitted to an exponential curve. All fitted values and their distributions are provided in the final output. For LSAE variance the mean is also provided.

8.2 An explanation of Model Maker results

When Model Maker completes analysis, select **Finish** to return to the Main Menu. Results are automatically saved in a results directory named in part MM. Open the results by selecting the Open Model Maker Directory icon:  in the upper right third of the Summary window.

Model Maker Report

DETAILS

STRMIX VERSION:	STRmix 2.7.0
USER:	jmorawitz
RUN DATE:	11 Jul 2019 12:08:51
REPORT RUN:	11 Jul 2019 12:20:33

MODEL MAKER SETUP

Single Source File	First 10 PROVEDIt MM 10rfu Genotypes Table.txt
Reference File	First 10 PROVEDIt MM 10rfu Genotypes Table_References.txt
Profiling Kit	GlobalFiler_3500_GS
Accepts	10,000
Seed	110378
Ignore peaks below detection threshold	N

PROBABILITY DISTRIBUTION

	α	β	MODE
Allele Variance c^2	3.772	4.293	11.900
Back Stutter Variance k^2	3.277	9.377	21.351
Forward Stutter Variance k^2	2.584	4.955	7.849
2bp stutter Variance k^2	1.821	5.595	4.593
Double back stutter Variance k^2	2.377	4.368	6.015
LSAE Variance	0.013		

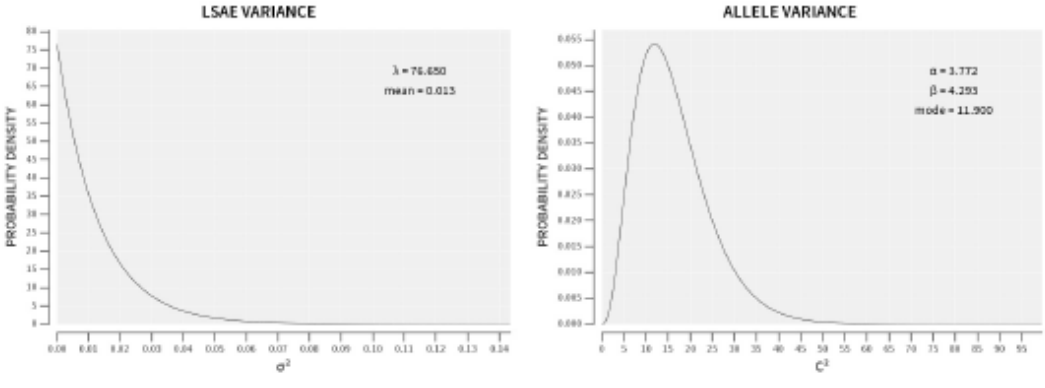


Figure 8.2 Example Model Maker results

The results include the following data:

1. STRmix™ version number, username and date of analysis, the names of the input files used, the kit used, the number of accepts and the seed used.

2. An indication of whether or not peaks below the analytical threshold that were found in the input files were ignored (this detail is also recorded in the log). This indication only appears if there were peaks present below the detection threshold in the input files.

3. α_x and β_x terms and mode for the fitted gamma curve for allele variance and stutter variance for each type of stutter modelled, and LSAE variance mean and plots

4. The number of profiles within the files and number of profiles used within the analysis (not shown in Figure 8.2)

5. A list of profiles not used within the analysis (not shown in Figure 8.2)

6. Correlation plots between LMW and HMW loci, and stutter and parent peaks (not shown in Figure 8.2)

7. Progression plot of log(likelihood values) over the analysis (Figure 8.3)

8. Summary of the settings and kit details used in the Model Maker analysis (not shown in Figure 8.2).

8.2.1 Log(likelihood) values

The log(likelihood) of the dataset at the current component-wise MCMC cycle is present 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. Figure 8.3 shows the log(likelihood) for one dataset over the course of the Model Maker MCMC run. In this dataset, the results plateau after approximately 30 iterations. The data can plateau either up or down.

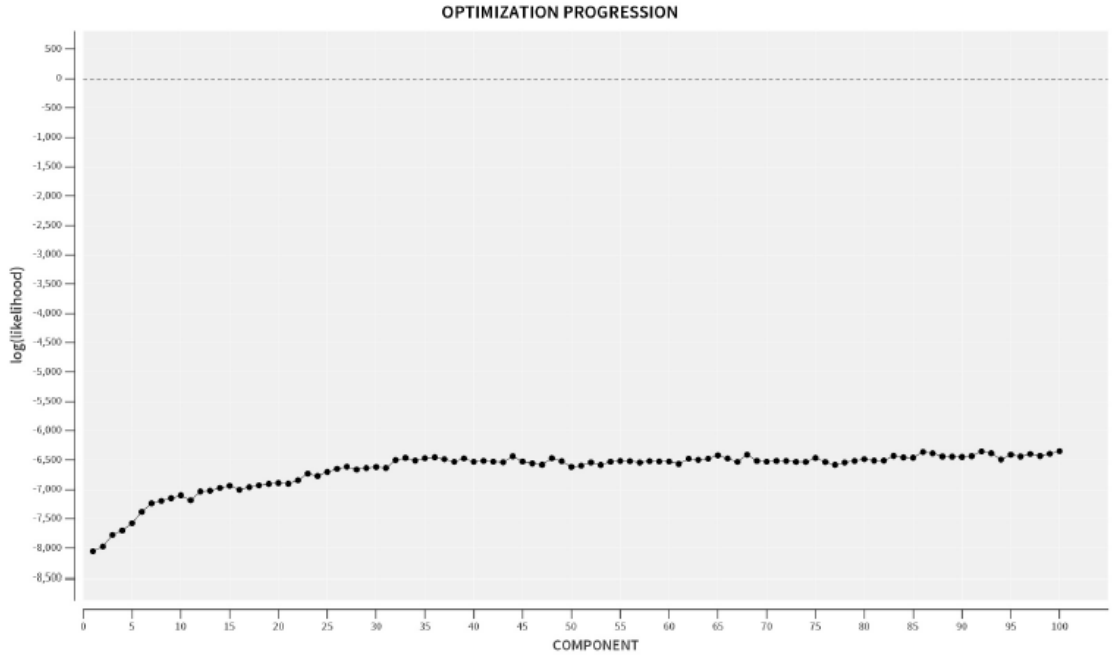


Figure 8.3 Plot of log(likelihood) values for Model Maker dataset over the course of the MCMC run

8.2.2 Correlation plots

The standard error of the log(*O*/*E*) values for the HMW and corresponding LMW allele within a heterozygote pair (allele correlation) and the standard error of the log(*O*/*E*) values for the allele and its corresponding stutter (stutter correlation) are plotted within the Model Maker report. These plots graphically show what level of correlation is present in the data (refer STRmix™ V2.8 Implementation and Validation Guide). Example plots are given in Figure 8.4.

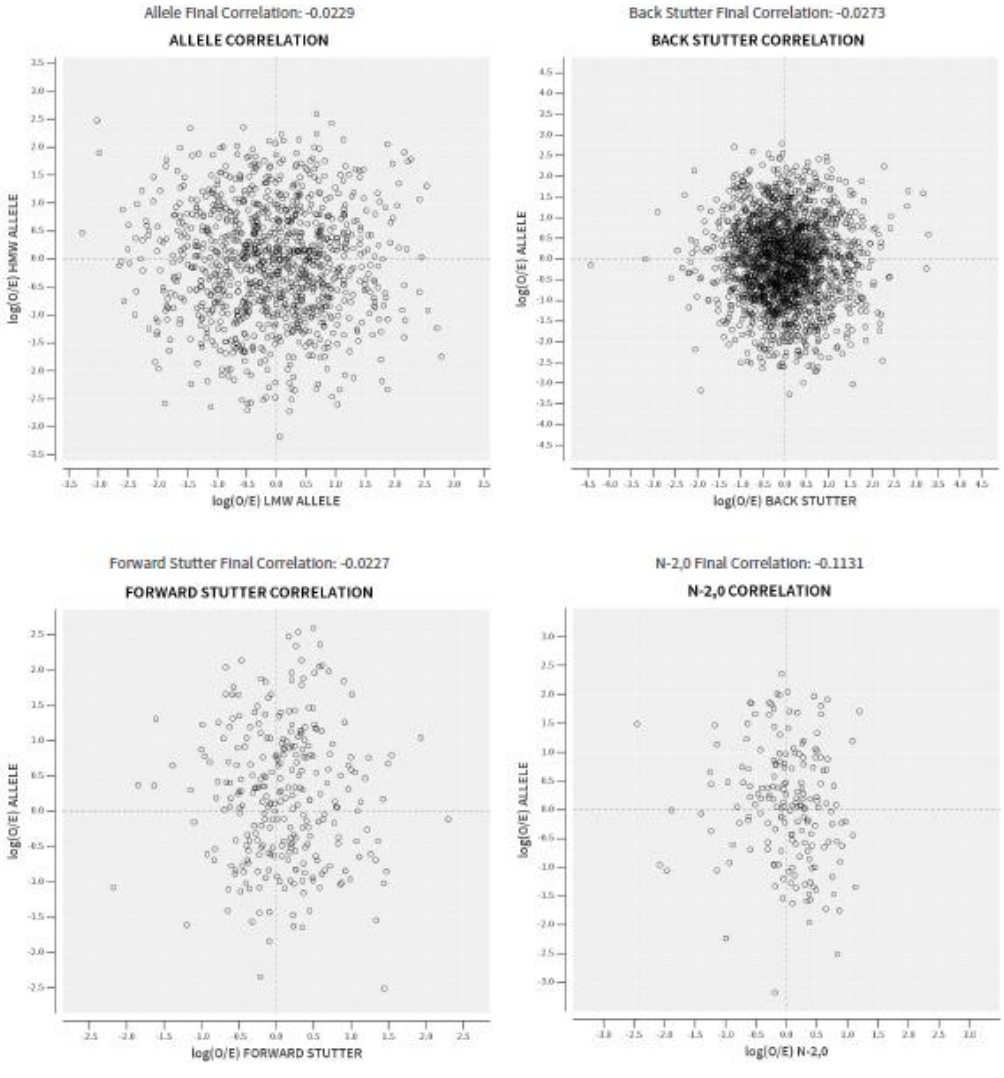
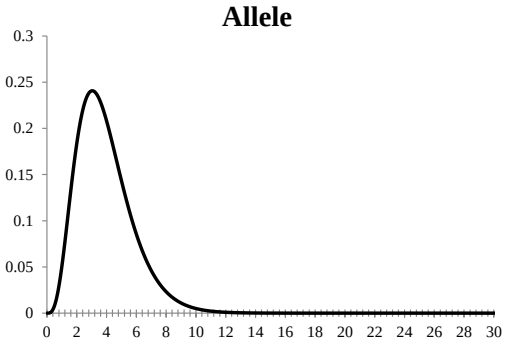
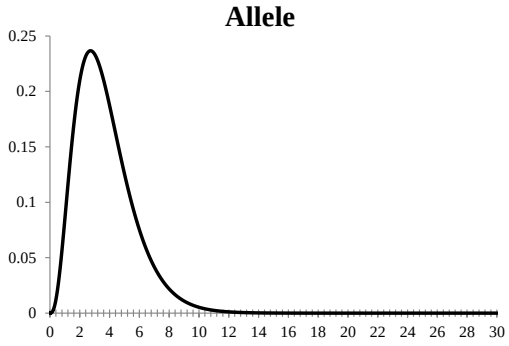
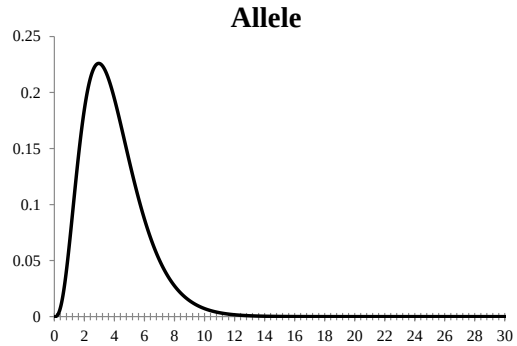
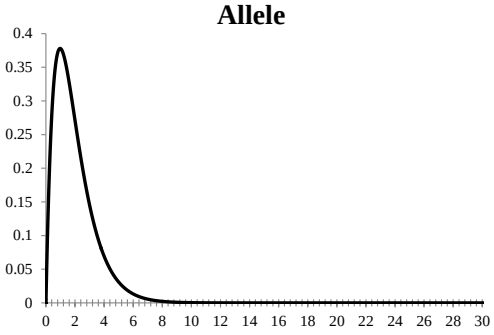
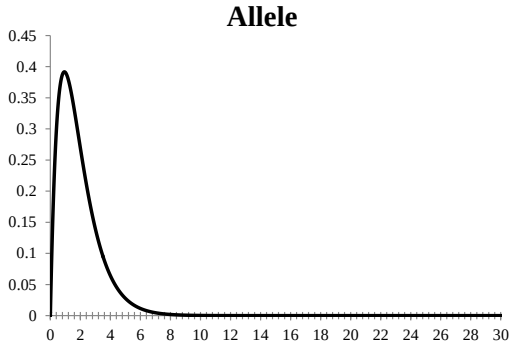
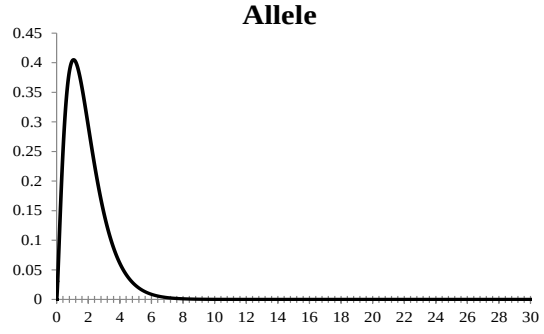
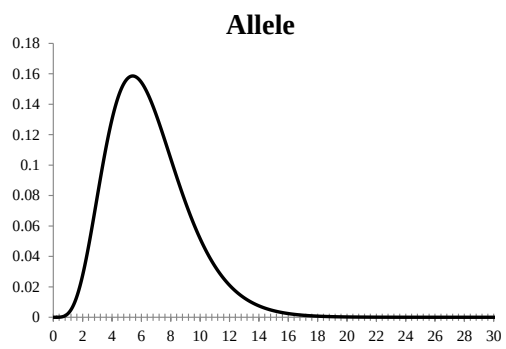
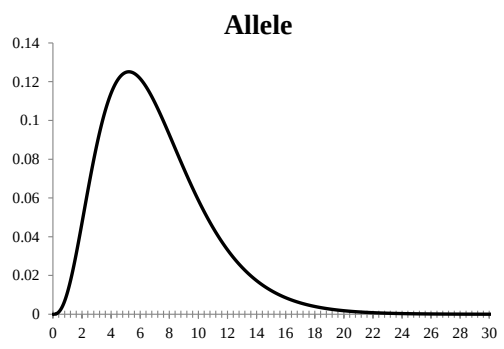
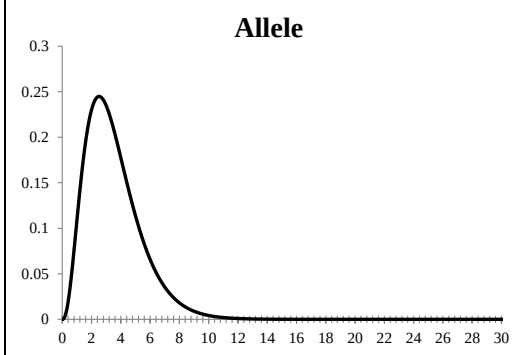
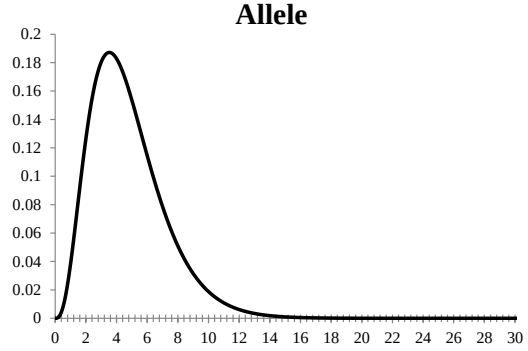
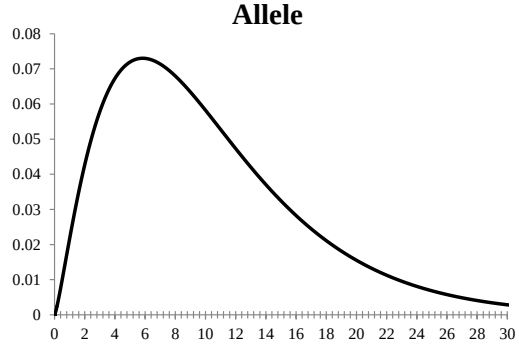
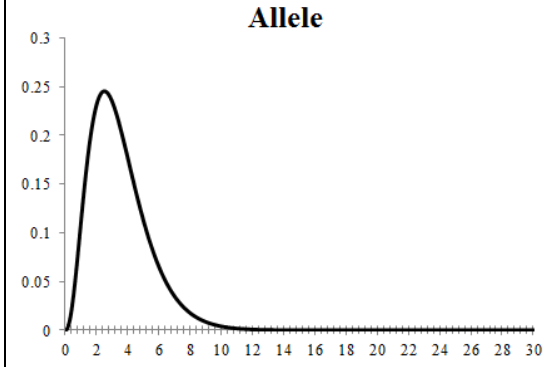


Figure 8.4 Allele, back stutter, forward stutter, and double-back stutter correlation plots from Model Maker report

8.2.3 Typical Model Maker results

Figure 8.5 shows the results of allele variance constant hyper-distributions after running Model Maker for 14 different datasets. Multiple regression of the mean of the gamma distributions produced in Figure 8.5 suggests that the largest factor in peak variability is PCR cycle number. CE instrumentation type (3500 versus 3130) and half versus full reaction volumes also had an effect on peak height variability. These results show that the differences between the same models of electrophoresis instrument using the same STR multiplex is minimal and so specific settings for each combination of hardware would not be required.

Set	*1*	2	*3*
Profiler Plus 28 cycles25μL	 Γ(4.53,0.862), mode 3.04, mean 3.9	 Γ (3.769,0.983), mode 2.72, mean 3.71	 Γ (3.962,0.997), mode 2.95, mean 3.95
Set	***4***	***5***	***6***
Profiler Plus 26 cycles 25μlL	 Γ(2.006,0.971), mode 0.98, mean 1.95	 Γ(1.994,0.942), mode 0.94, mean 1.88	 Γ (2.339,0.801), mode 1.07, mean 1.87

Set	7	8	9
29 cycles	GlobalFiler™ 25 µL	PowerPlex 21 12.5 µL	PowerPlex 21 25 µL
			
	Γ(5.78,1.131), mode 5.41, mean 6.54	Γ(3.84,1.838), mode 5.22, mean 7.06	Γ (3.54,0.99), mode 2.51, mean 3.50
Set	**10**	**11**	**12**
	GlobalFiler™ - 29 cycles – 3130xl 25µL	GlobalFiler™ - 29 cycles – 3500 25µL	Identifiler® - 28 cycles 25 µL
			
	Γ(3.907,1.215), mode 3.53, mean 4.75	Γ(2.295,4.507), mode 5.84, mean 10.34	Γ (3.569,0.982) , mode 2.52, mean 3.50

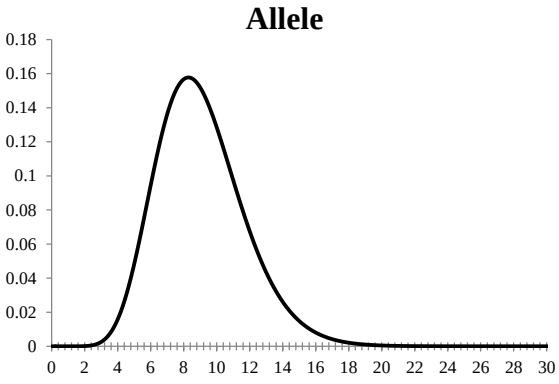
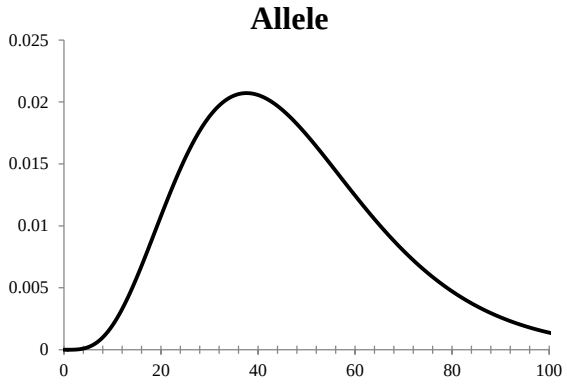
Set	13	14
	MiniFiler™ - 30 cycles – 50 µL	SGMPlus - 34 cycles – 50 µL
		
	Γ (11.906,0.760), mode 8.29, mean 9.05	Γ (4.983,9.449), mode 37.64, mean 47.08

Figure 8.5 Comparison of peak variance between different data sets run through Model Maker

Some typical Model Maker results for a number of STR profile kits, cycle conditions and instrumentation are provided in Table 8-1 below, where the gamma distribution of the variance constants are described by their parameters α,β and their mode $(\alpha-1)\beta$.

Validated kits from STRmix™ version 2.4, 2.5, and 2.6 are compatible with STRmix™ 2.7. Forward stutter variances will automatically be populated with the allele variance constant (if using kits from versions 2.4 and 2.5), as this was the variance constant used for forward stutter in these versions of STRmix™. However, it is highly recommended that Model Maker is run again in STRmix™ 2.7, and a performance check is required.

Table 8-1 Example of different variance constant distributions for different platforms

Kit	Cycle Number	CE	Allele Variance (mode)	Back Stutter Variance k^2 (mode)	Forward Stutter k^2 (mode)	(0,-2) Stutter k^2 (mode)	(-2,0) Stutter k^2 (mode)	LSAE mean
Fusion 5C	29	3500	3.351,7.044 (16.558)	1.521,18.307 (9.53)	2.579,3.467 (5.473)	1.605,2.408 (1.456)	1.57,9.257 (5.279)	0.015
			3.54,6.697 (17.011)	1.543,17.705 (9.605)	3.587,2.373 (6.138)	1.798,1.838 (1.467)	1.507,11.551 (5.862)	0.016
			3.537,7.305 (18.532)	1.51,17.793 (9.079)	4.276,2.018 (6.612)	2.333,1.44 (1.92)	1.56,10.391 (5.823)	0.015
			5.352,4.258 (18.533)	1.542,18.238 (9.891)	2.953,2.82 (5.507)	1.673,2.293 (1.543)	1.563,11.285 (6.353)	0.014
			4.262,5.339 (17.415)	1.529,18.266 (9.66)	3.653,2.342 (6.213)	2.832,1.504 (2.756)	1.529,11.275 (5.966)	0.016
Fusion 6C	29	3500	6.468,1.097 (5.999)	1.525,11.448 (6.012)	2.423,3.689 (5.25)	1.818,2.968 (2.426)	2.761,2.396 (4.221)	0.01
			7.539,0.929 (6.076)	1.518,11.954 (6.191)	2.157,4.265 (4.937)	1.57,3.933 (2.24)	2.233,3.024 (3.728)	0.01
			6.26,1.144 (6.019)	1.521,11.832 (6.168)	1.971,4.792 (4.652)	1.504,4.563 (2.301)	2.657,2.393 (3.966)	0.009
			6.424,1.139 (6.178)	1.55,10.876 (5.982)	2.057,4.604 (4.867)	1.575,4.164 (2.396)	2.567,2.562 (4.014)	0.008
			7.413,0.937 (6.012)	1.533,11.597 (6.184)	2.251,4.383 (5.484)	1.538,3.767 (2.026)	2.594,2.369 (3.777)	0.011
GlobalFiler	29	3500	3.505,5.49 (13.754)	1.513,10.877 (5.575)	1.518,17.18 (8.9)	1.559,3.29 (1.84)	7.154,2.173 (13.372)	0.037
			10.466,1.693 (16.022)	1.51,10.812 (5.511)	1.644,14.932 (9.617)	1.835,2.492 (2.081)	6.214,2.522 (13.152)	0.029
			2.472,8.62 (12.686)	1.506,10.772 (5.448)	1.729,15.028 (10.961)	1.66,2.796 (1.846)	5.935,2.717 (13.406)	0.029
			7.019,2.599 (15.641)	1.522,10.371 (5.417)	1.525,15.718 (8.253)	1.621,2.489 (1.546)	4.493,3.477 (12.143)	0.028
			6.103,2.843 (14.511)	1.506,10.733 (5.428)	1.504,16.494 (8.32)	1.588,3.514 (2.068)	6.443,2.691 (14.647)	0.025
Investigator	29	3500	3.595,0.93 (2.413)	1.504,4.52 (2.278)	2.24,4.462 (5.535)	2.029,1.364 (1.404)	2.802,6.629 (11.943)	0.008
			3.309,0.987 (2.279)	1.511,4.78 (2.442)	2.106,4.929 (5.452)	2.341,1.199 (1.608)	3.023,5.665 (11.464)	0.008
			3.404,0.939 (2.258)	1.51,4.575 (2.331)	2.233,4.767 (5.879)	1.873,1.492 (1.302)	3.212,5.878 (12.999)	0.007

9. AUDIT FILES AND HASHES

The SHA-512 cryptographic hash is the 512-bit component of the “SHA-2” data integrity check standard. SHA-512 generates an almost-unique 512-bit (32-byte) digital signature or fingerprint for a text. These hashes are generated for the input and results files in each STRmix™ interpretation. The hash results for the inputs, configuration files (config) and results are printed at the end of the pdf report (see Figure 9.2). To disable this option, please refer to sections 6 and 7.5 above)

Hash results are found in the Audit folder (see Figure 9.1). Hashes can be used to help prove that no tampering with the results has occurred. A comparison of the hash results found within the Audit folder will match the hash results printed to the pdf report if neither the pdf report or hash files have been altered or tampered with. Re-running the calculation with exactly the same conditions should also generate the same hashes between runs.

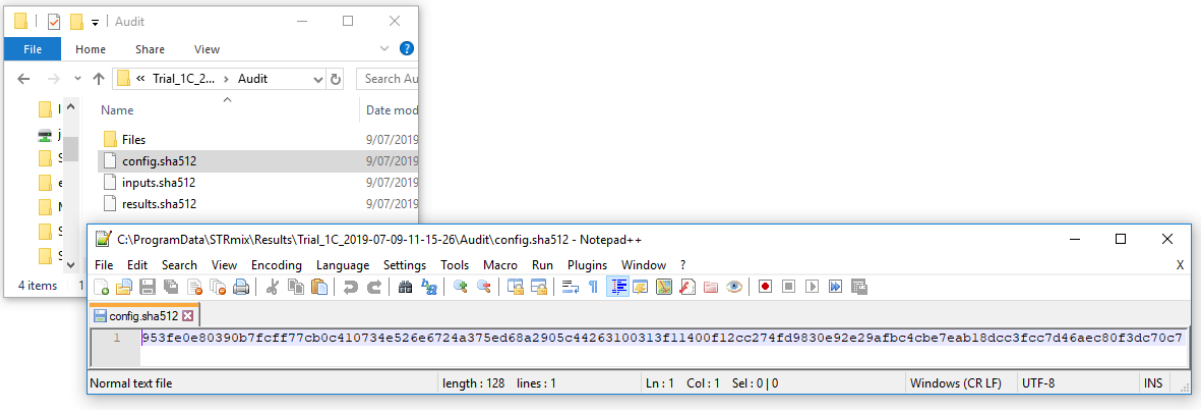


Figure 9.1 Example of the hash files found within the Audit folder of the Results

AUDITING DETAILS

SHA-512 checksum Information

Configuration	953fe0e80390b7fcff77cb0c410734e526e6724a375ed68a2905c44263100313f11400f12cc274fd9830e92e29afbc4cbe7eab18dcc3fcc7d46aec80f3dc70c7
Inputs	334c59613ad57d16bf345c4a63a0c762ce70685e593538e964dcc3927c0d3af251debbeb24f1c989053a9c850078ec326aa6b4483814bb73f00be2f3460ec882
Results	285db73aa88d8a613625bf2b1c0f890711a4e4fc876b992467d6013d26e7ef156ac6c0fe2ca9720d8a9b83425d9b361be5bdc642cc316049388a49f5ad026b0b

Figure 9.2 Example of the Auditing Details found at the end of the pdf report

The data used to generate the config, inputs, and results SHA-512 hashes is a select combination of the various essential elements that go into each of these (see Table 9-1 for detail). For example, a hash simply generated from the config.xml file will not match the config.sha512 hash because it excludes some details.

The Files folder within the Audit folder contains SHA-512 hashes of each file in the results folder. If uploading Results folders to a LIMS system, the hash values within this folder can be checked using a data hashing tool to confirm that the data has not been corrupted on transfer and that the correct files are present within the Results folder.

In STRmix™ V2.8, a new hash file has been added that includes only the input files used in the run and does not include the kit information. This hash is called the samples.sha512 hash. The inputs hash (inputs.sha512) remains and still contains the input files as well as the kit used in the run.

Table 9-1 Table of Audit Hash file contents

Hash	Contents
Config.sha512	Config.xml, excluding the Case Number, SampleID and Comments fields
Inputs.sha512	Kit, stutter, exception and regression files. Sample input files, population and allele frequency files.
Results.sha512	Results.xml (excluding the Run Info), GenotypePDF and Component Interpretation files.
Samples.sha512	Input evidence and reference files only.

10. RELEASE NOTES

10.1 V2.7 to V2.8 Summary of changes

10.1.1 General use

- Improvements to both low and normal memory modes to allow for processing of larger problems with lower RAM requirements
- Code refinements
- Improvements to the packaging of the STRmix™ software
- Improvements to logging
- Improvements to when warning messages regarding ignoring peaks below AT are shown
- Continued improvements to error message handling of null pointer and uncaught errors
- Fix to a compatibility issue with Non-English (French) date format preventing a v2.5 LR from previous running in v2.6 and 2.7
- Fix to compatibility issue with accents (eg é) used in comments
- Addition of VeriFiler™ Plus kit type
- Various improvements to aid in developmental testing and validation
- Improvements in internationalisation to handle grammatical gender usage
- Change to Model Maker to check for partial references against the kit itself rather than against the evidence input samples
- Improvements to tickbox setting display in the User Interface
- Improvement to allow loci to be ignored in LR from Previous
- Allow batch mode to exit on finish (for export from FaSTR™ DNA)
- Changes to the way retrospective drop-in alleles are calculated for investigations on runs in versions prior to STRmix™ v2.5 where this information was not written to the results folder
- ‘Drop-in frequency’ in user interface changed to ‘Drop-in rate parameter’

- TM added to STRmix files on disc
- Updates to website links in About screen, installer and Purchase button
- Improvements to Auto Database search so that it is prohibited if the setup is fully conditioned
- Update to the installer background image website
- Fix to display milliseconds correctly in "human readable" time formatting when execution time is over a day
- Change to allow Allele frequency files and stutter exceptions files not to contain non-relevant (gender/quality/Y/ignored) loci
- Updates to the Software Licence Agreements

10.1.2 Changes to science

- Improvement to the model switch at quarter burn-in
- Improvements to modelling of drop-in peaks in stutter positions
- Template output within report PDF changed from mode to mean
- Normalisation factor included in *LR* extended output
- Introduction of Minimum Resampled Count in the Populations to allow control of the minimum allele frequency separately from the population size within the HPD calculation
- Model improvements in allele frequency sampling to use k+1 (instead of k)
- Change for total iterations counts from integers to longs to prevent the possibility of too many iterations causing an integer overflow
- Change to automatically ignore loci that are missing from a partial reference when used for conditioning
- Change to include Q allele resampling during HPD
- Change to include unobserved alleles in the allele frequency normalization within HPD
- Improvements to template sampling for multi-kit interpretations
- Change within the HPD to use the k value for the unobserved allele from the current locus (not from locus 1)
- Change to allow starting DNA amounts to align with Mx priors in Smart Start
- Improvements to ESS thinning
- Improvement to allow weights to be resampled in HPD iteration 1
- Change to Linear approximation for degradation to use mean (instead of mode) for consistency

10.1.3 New Features

- Addition of a Top Down Approach to mixture interpretation to allow an *LR* to be generated for only the nominated major contributors to a profile
- Ability to use a database file as a reference input file

10.1.4 Changes to reports

- Fix to resolve an issue where not all STRmix™ v2.6 reports were able to be compiled in later versions of STRmix™
- Internationalisation of reports
- Removal of Calculation Array Check from extended outputs

- Improvements to Model Maker extended output file including adding profile indices and filenames
- Removal of run information from Results.sha512 hash to allow the use of the Results.sha512 as an indicator of same results between two runs
- Addition of a new audit hash (Samples.sha512) to include input files only
- Addition of the Minimum Resampled Count per population to PDF report
- Inclusion of count of log(*LR*)=0 on the H_d True Tester plots in the report
- Addition of assumed partial loci to the PDF report when conditioning on a partial reference
- Improvement to *LR* from Previous report: When running an *LR* from Previous, if the seed from the interpretation cannot be found, rather than listing the *LR* as 0, this line is no longer populated in the PDF report
- Addition of inter-section spacing before the last report component in pdf reports
- Replacement of 3.2 with 3.3 in CODIS cmf report
- Inclusion of genotype sets and re-calculated profile probability from burn-in to the Post Burn-in extended output file
  Improvement to allow the *LR* sorting in the PDF report for Familial searches to handle a blank value

11. REFERENCES

[1] S. Gittelson, T. Kalafut, S. Myers, D. Taylor, T. Hicks, F. Taroni, I.W. Evett, J.A. Bright, J. Buckleton, A Practical Guide for the Formulation of Propositions in the Bayesian Approach to DNA Evidence Interpretation in an Adversarial Environment, Journal of Forensic Sciences 61(1) (2016) 186-195.
[2] 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.
[3] K. Slooten, Identifying common donors in DNA mixtures, with applications to database searches, Forensic Science International: Genetics 26 (2017) 40-47.
[4] 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.
[5] J.-A. Bright, D. Taylor, Z. Kerr, J. Buckleton, M. Kruijver, The efficacy of DNA mixture to mixture matching, Forensic Science International: Genetics 41 (2019) 64-71.
[6] M. Kruijver, J.-A. Bright, H. Kelly, J. Buckleton, Exploring the probative value of mixed DNA profiles, Forensic Science International: Genetics 41 (2019) 1-10.
[7] C. Brookes, J.-A. Bright, S. Harbison, J. Buckleton, Characterising stutter in forensic STR multiplexes, Forensic Science International: Genetics 6(1) (2012) 58-63.
[8] J.-A. Bright, D. 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.