STRmix™ Probabilistic Genotyping Software Operating Instructions

This procedure describes the use of STRmix™ V2.4 for the interpretation of PowerPlex® Fusion DNA profiles within the NYC Department of Forensic Biology. Readers are also referred to the STRmix™ v.2.4 Users and Operation manuals for additional information.

I. Preparing Data for a STRmix™ Analysis

A. Before performing your STRmix™ analysis, the following actions must be taken:

1. Verify that the sample is suitable for STRmix™ analysis (Refer to the STR Results Interpretation manual)

2. Determine the best described Number of Contributors to the sample (NOC). Refer to the STR Results Interpretation manual regarding the procedure for determining the number of contributors.

3. Create folders for the STRmix™ runs:

   a. Navigate to the M:\STR_Data\STRmix Data Folder

   b. Within the STRmix Data folder, create a new folder with the FB (or FBS) case number

e.g. FB16-01234 or FBS16-05678

   c. Within the FB (or FBS) case file folder, create a folder for EACH evidence or suspect sample that will be run through STRmix™. Use the sample's OCME ID for the naming of the folder. Suffixes such as 'mcon' or 'reamp' should not be included.

e.g. FB16-01234_567_1_1.1_trig_GS
FB16-01234_890_1_1.1_shirt_BL
FB16-01234_123_1_1.1_VS_SF
FBS16-05678_999_1_1.1_(s)JS
FBS16-05678_888_1_1.1_cupJS

Note: If a suspect is being compared to multiple FB's, create a sub-folder within that suspect sample folder for each cross-referenced FB.

4. Confirm that the STR data is prepared correctly for STRmix™ analysis:

   a. Evidence samples must be edited to remove all artifacts, including pull ups, spikes, dye blobs, n-8 stutter and n-2 stutter, etc before inputting into STRmix™. Standard forward and reverse stutters must ***not*** be removed before importing into STRmix™

   b. Evidence samples must only be amplified in PowerPlex Fusion® in order to undergo STRmix™ analysis and an attempt should be made to amplify conditioned and comparison samples in PowerPlex Fusion® as well. If being utilized, comparison and conditioned samples must be edited to remove all artifacts and all stutter.

   c. STRmix™ allows the user to calculate a likelihood ratio when the evidence and conditioned /comparison samples are analyzed in different autosomal typing kits. LRs will only be provided for those loci in common between the two kits.

      i. In the case of an evidence and conditioned/comparison sample amplified using different autosomal kits, both the evidence and the conditioned/comparison samples have to input from .txt files and the locus order must match that of the evidence (PowerPlex Fusion® order). The comparison/conditioned sample data can be converted to the proper PowerPlex Fusion® order and appropriate .txt file format using the following macros:

For suspect, victim, elimination samples run in Identifiler use macro:
"Identifiler to Fusion Exemplar STRmix Input Creation".

For deconvoluted DNA Donors, use macro:
"Fusion Deconvoluted STRmix Input Creation"

   d. Non numeric values such as OL or OB, < 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

i. Assignment of an actual allelic size designation to a non numeric value within the input file:

a) Open the STRmix .txt file associated with the appropriate STR project (e.g. in Notepad)

b) Locate the sample and locus containing the non numeric value within the .txt file and manually replace the value with the appropriate actual allelic size designation

c) Save .txt file replacing the original file

e. Locations that contain a tri-allelic pattern or have an unresolved peak must be ignored for proper STRmix™ analysis.

i. Ignoring a tri-allele locus in a single source evidence sample during STRmix™ deconvolution:

a) In **Configure Analysis** window click **Run Settings** at the bottom of the window.

b) Under **Thresholds** settings click **Ignore Loci** button.

c) Select the locus within the **Include** window and click the > button to move the locus to the **Exclude** window.

d) Click save when all appropriate loci have been added to the **Exclude** window.

e) Click save again in the **Run Settings** window and continue with instructions for performing STRmix™ deconvolution below in section **II.**

ii. Ignoring a locus with unresolved peaks during LR from previous analysis:

a) In **Configure Analysis** window click **Run Settings** at the bottom of the window.

b) Under **Thresholds** settings click **Ignore Loci** button.

c) Select the locus within the **Include** window and click the > button to move the locus to the **Exclude** window.

d) Click save when all appropriate loci have been added to the **Exclude** window.

e) Click save again in the **Run Settings** window and continue with instructions for performing LR from previous analysis below in section **III.**

f. Sample data must be assembled into the appropriate format for STRmix™ input. The standard input for STRmix™ are .txt files . See "STRmix™ analysis for Evidence" and "Exporting Exemplar Table for STRmix™ input" sections within the GeneMarker manual.

g. Prepare the scenarios to be run in STRmix™.

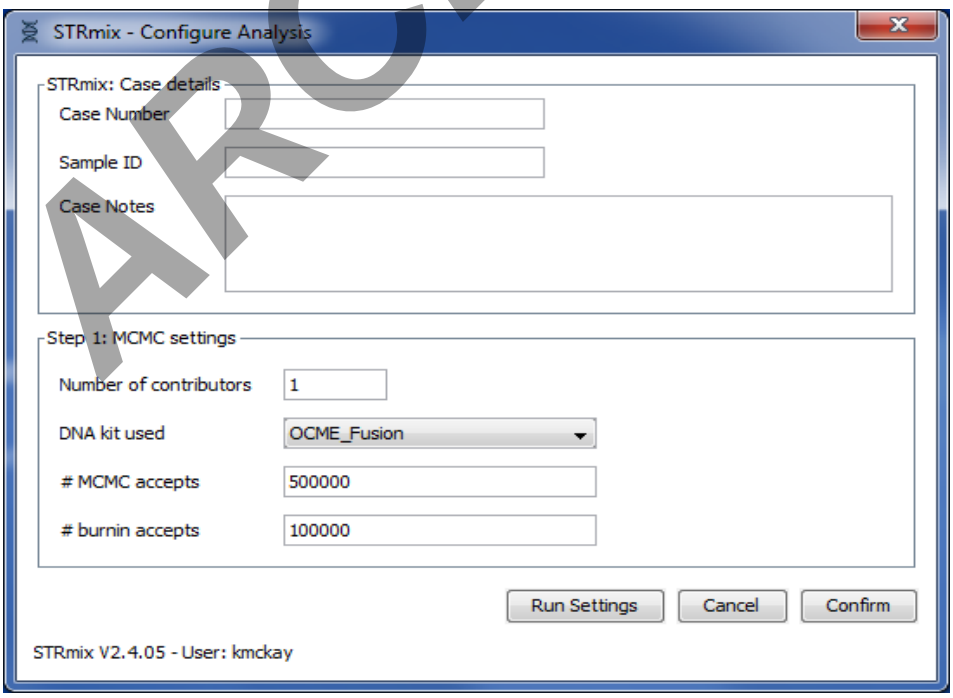
i. Are you performing a deconvolution on an evidence sample? Go to Section II: **Deconvolutions in STRmix™**

ii. Are you performing STRmix™ to evaluate a comparison sample against an evidence sample and generate an LR, with or without conditioned profiles? Go to Section III: **Likelihood Ratio calculations with STRmix™.**

II. Deconvolutions in STRmix™.

A. Launch the STRmix™ application. Open the STRmix™ software by locating STRmix™ in the task bar or by double clicking on the STRmix™ icon on the desktop.

B. Select "Start Analysis" from the startup screen. This will open the "STRmix – Configure Analysis" window.



C. Naming STRmix™ runs

STRmix™ output folder and file names are created by stringing together the values entered into the "Case Number" and "Sample ID" fields in the software followed by the date and time of the analysis run. The information in the file name is separated by dashes. Therefore, if other characters are entered, such as a comma, underscore, period, etc., the software will convert them into dashes.

Use the following naming convention for deconvolutions:

Case Number = YY-XXXXX (leave out "FB")
Sample ID = remainder of the OCME ID* #p (NOC) cond elim initials (if applicable)
Case Notes = leave blank

*Suffixes such as 'mcon' or 'reamp' should not be included in the OCME ID.

e.g. deconvolution of 3 person mixture, FB16-01234_567_1_1.1_trig_GS, conditioned on John Doe

Case Number = 16-01234
Sample ID = 567_1_1.1_trig_GS 3p condJD

Examples	Resulting STRmix file name
1-person deconvolution	16-01234-567-1-1-1-trig-GS-1p
2-person deconvolution, no conditioning	16-01234-567-1-1-1-trig-GS-2p
2-person deconvolution, conditioning victim AB	16-01234-567-1-1-1-trig-GS-2p-condAB
3-person deconvolution, no conditioning	16-01234-567-1-1-1-trig-GS-3p
3-person deconvolution, conditioning elim CD	16-01234-567-1-1-1-trig-GS-3p-condCD
3-person deconvolution, conditioning elims CD and EF	16-01234-567-1-1-1-trig-GS-3p-condCD EF

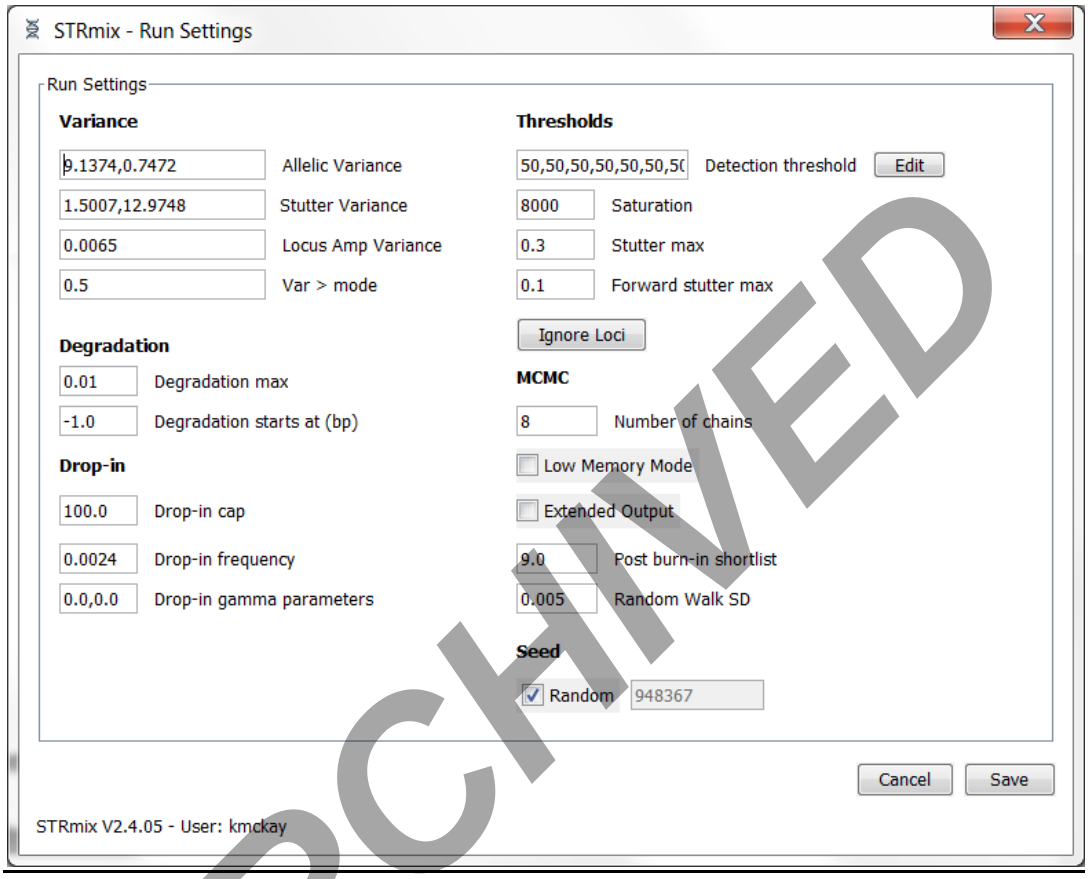
D. Set the **Number of Contributors.**

1. Ensure that the following "Step 1: **MCMC settings"** are in place:

DNA kit used: OCME_Fusion
 # MCMC accepts: 500000
 # burnin accepts: 100000

NOTE: the # MCMC accepts and # burnin accepts must not be modified without documented approval from the technical leader.

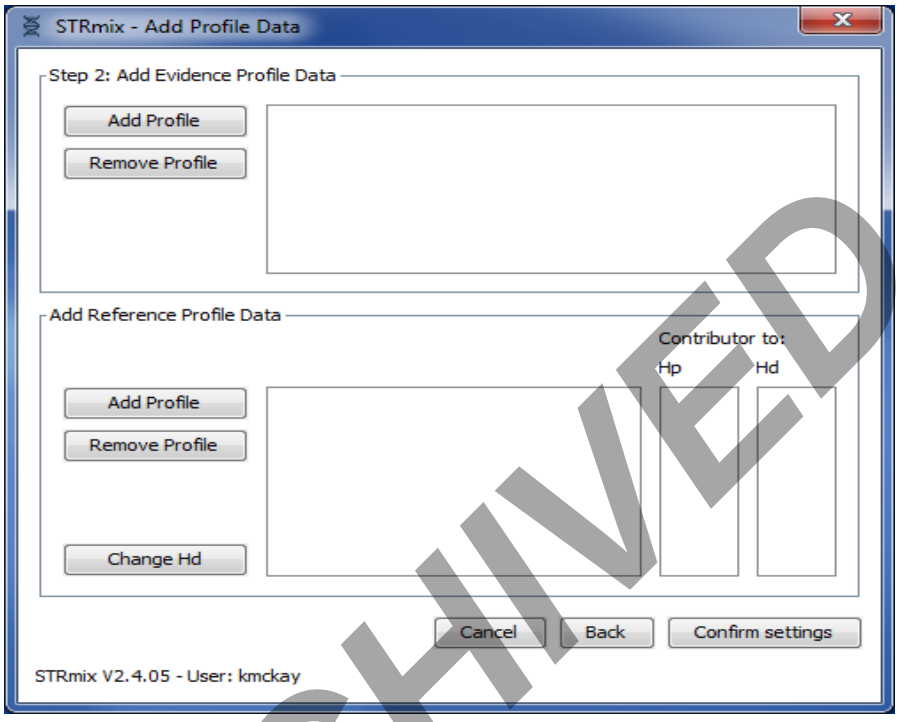
2. Select "run settings" to confirm run settings. The settings should be as follows for every STRmix™ analysis. Any changes that are made will appear in bold on the run report. Press "Cancel" when done. See section **I.A.4.e** above for instructions on ignoring loci under specific situations.



3. Select "Confirm" to proceed to the "Add Profile Data" window, or cancel to return to the Startup screen (canceling will not save any data up to this point). **Once "Confirm" is hit, a folder will be created in the STRmix results folder on your C drive. If incorrect nomenclature was used and you return to this screen to make changes, the empty folder on your STRmix results drive should be deleted.**

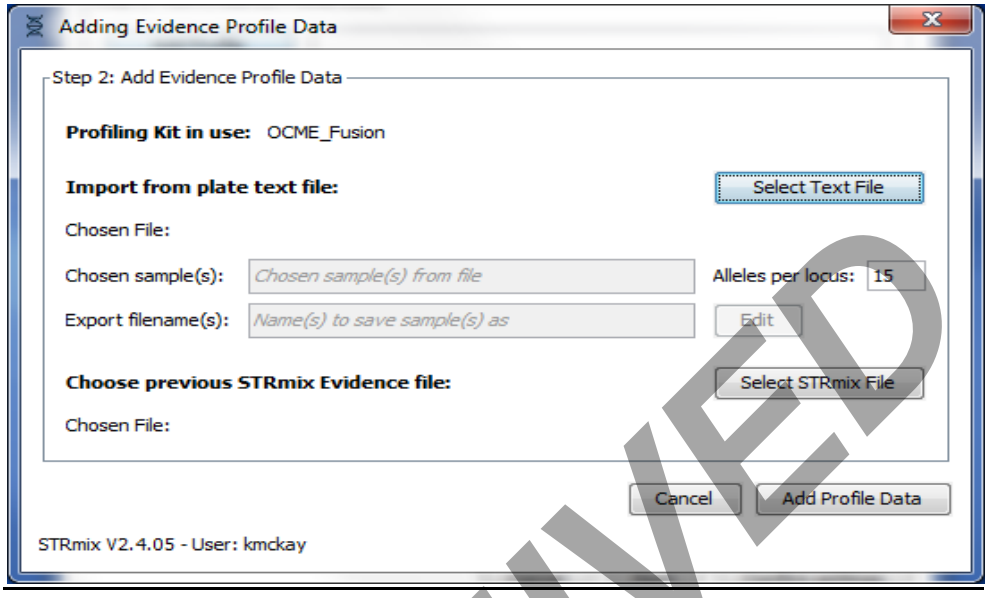
E. Add your Evidence Profile Data.

1. In the "STRmix – Add Profile Data" window, select "Add Profile". This will bring you to the "Adding Evidence Profile Data" window (see step 2 below) where you can select either a text file or STRmix file from which to run a STRmix™ analysis. Alternatively, you can navigate to your data folder and drag and drop the appropriate text file into the top box, and proceed to step 4.

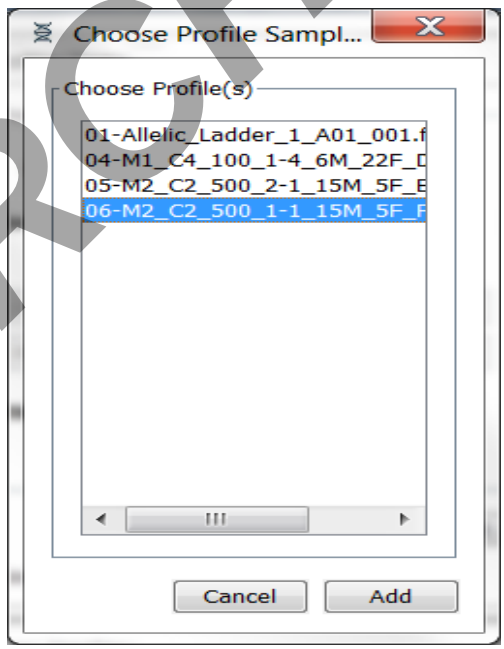


2. In the **"Adding Evidence Profile Data"** window (see below), ensure that **"Alleles per locus"** is set to 15.

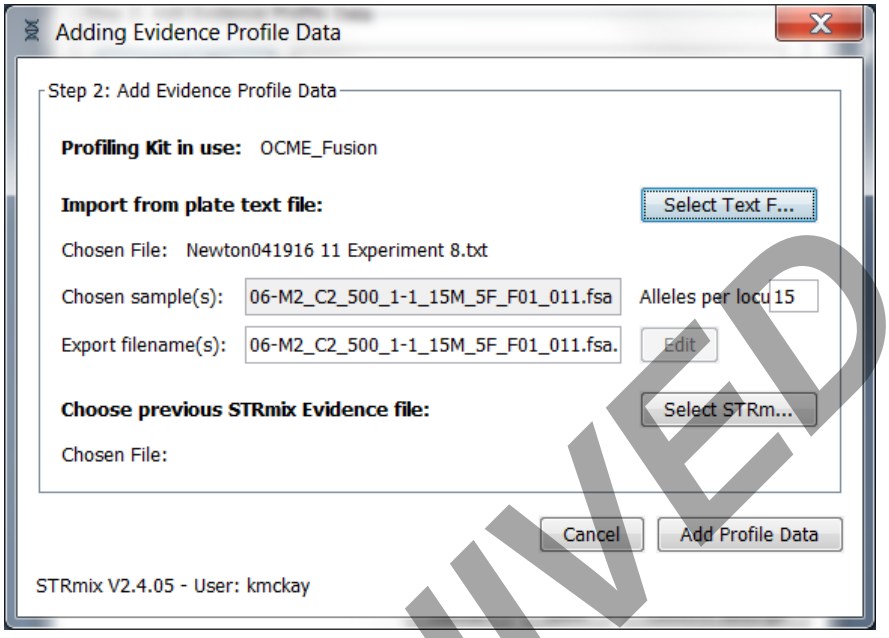
3. Choose "Select Text File" to navigate to and import your STRmix STR run data from the STR data folder.



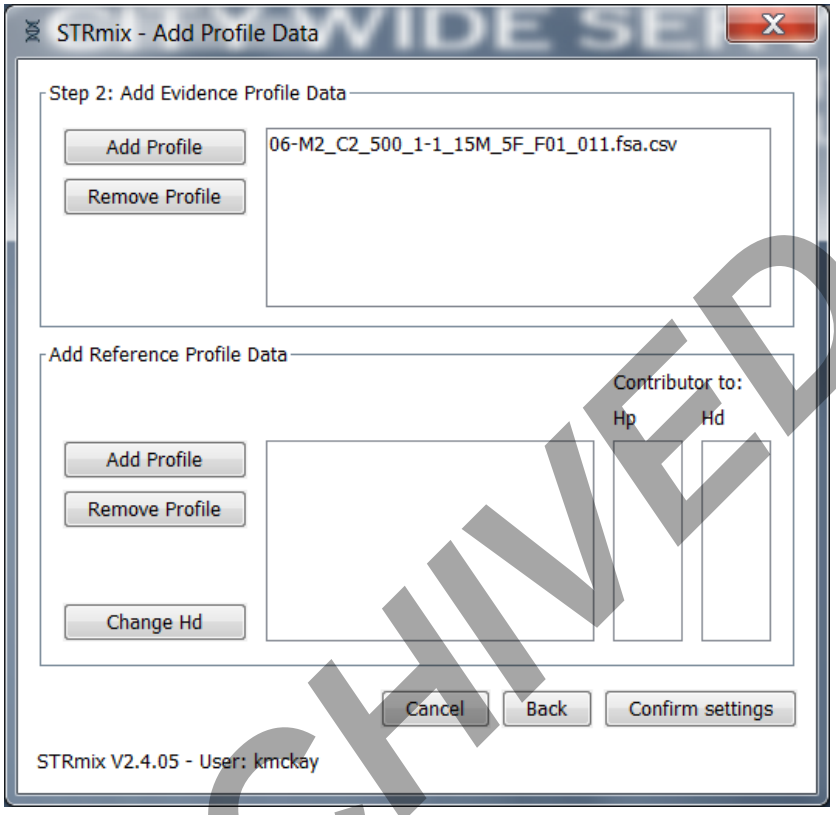
4. When you select a text file with multiple samples, it will look as follows:



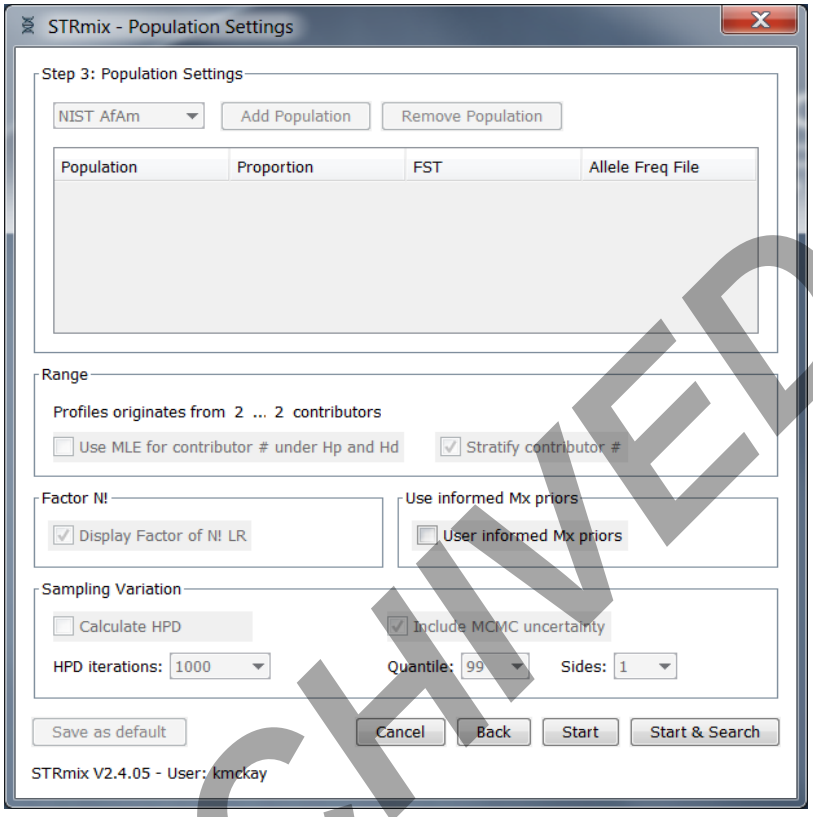
5. Select the appropriate sample by clicking on it, then select "Add" for the relevant text file. Your view will appear as follows:



6. Now select "Add Profile Data" and your text file name will appear in the "Add profile Data" screen as seen below. **Repeat steps 1-6 to add any replicates of the sample which were amplified.**



F. **Deconvolutions without a conditioned contributor**. Select "Confirm settings" and this will open up the "Population Settings" window. For a deconvolution without a conditioned contributor, population data is not needed, therefore the populations will appear grayed out in the screen below. Select "Start" to begin your analysis. Proceed to Step H.



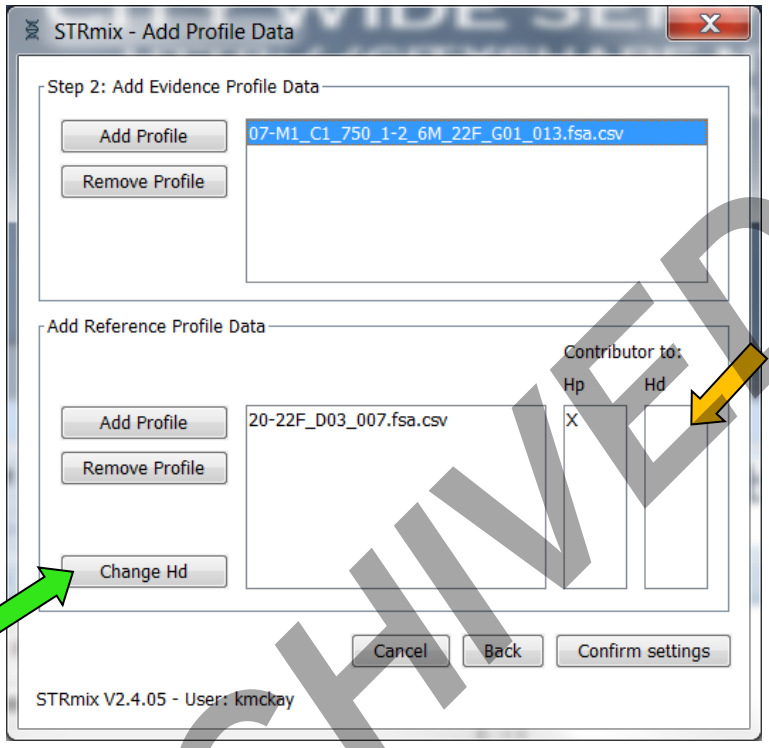
G. **Deconvolutions with a conditioned contributor:** For deconvolutions with a conditioned contributor, you must also add reference profile data. Refer to the STR Results Interpretation Manual for guidance on when a conditioned contributor may be applied. A deconvolution of the evidence sample without conditioning and a LR against a potential conditioned contributor may need to be run first.

1. Select "Add profile" under the "Add Reference Profile Data" section of the screen, or drag and drop the file in to the reference sample box and skip to step 3.

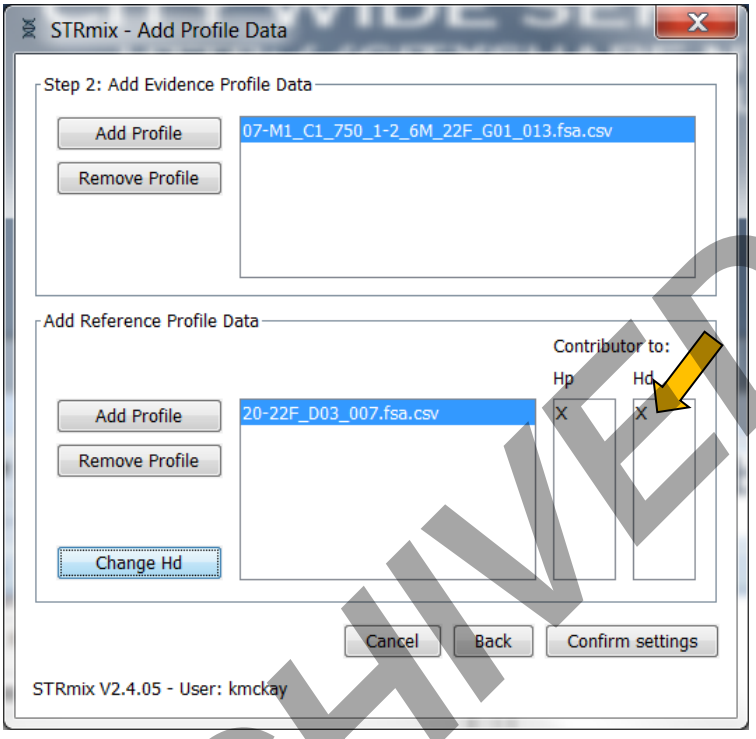
2. Click on "Select Text F..." to navigate to the case conditioned reference text file.

3. Once you open up the reference text file, click on it and select "Add". This will bring you to the following "Adding Reference Profile Data" screen. Select "Add Profile Data" to complete the process of adding your conditioned contributor's profile to the analysis.

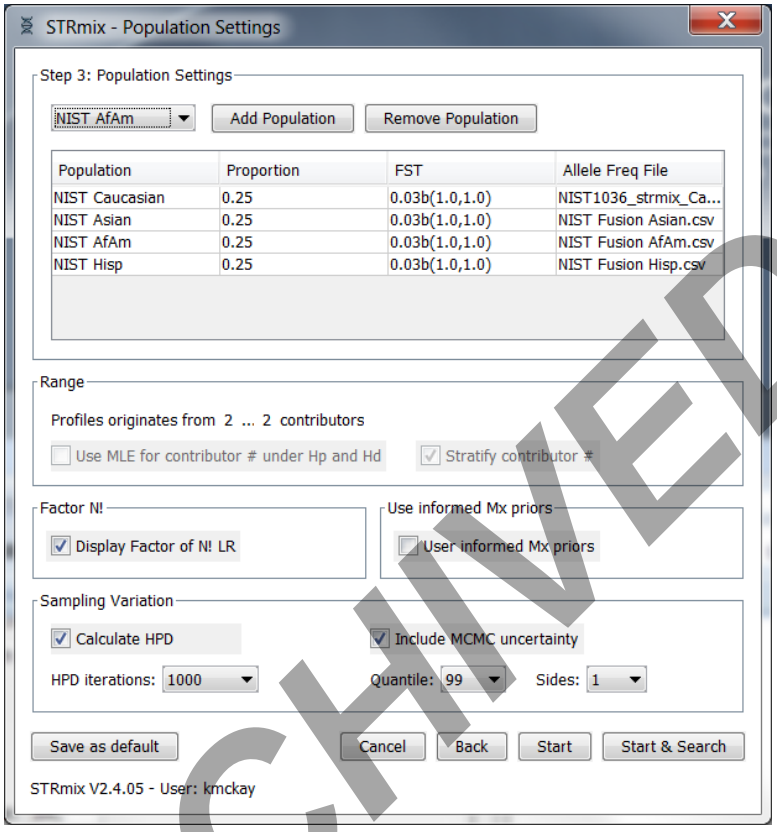
4. The conditioned sample will now appear in the "Add Reference Profile Data" section of the screen below.



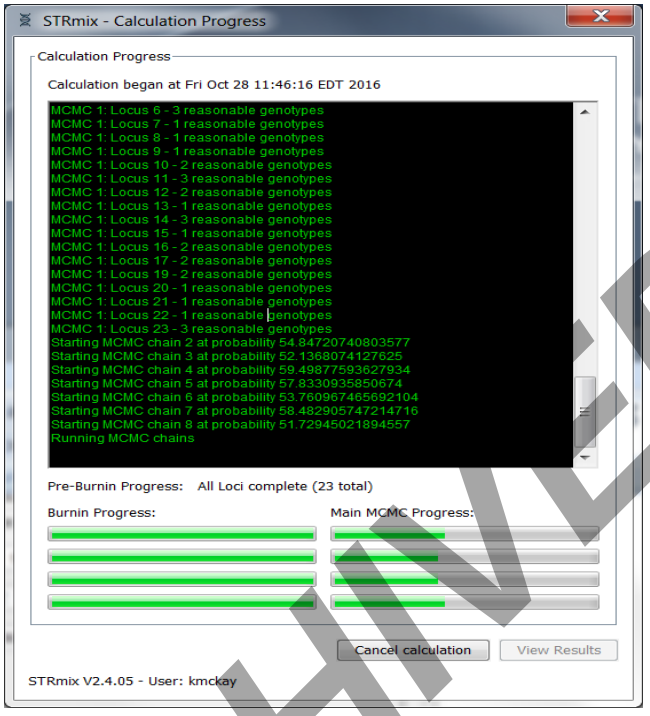
5. Conditioned contributors are considered true donors in Hp and Hd. Therefore, you must assign the conditioned contributors as such by selecting "Change Hd". This will allow the conditioned contributor to be chosen under H_p and H_d and will have an "X" marked in both columns.



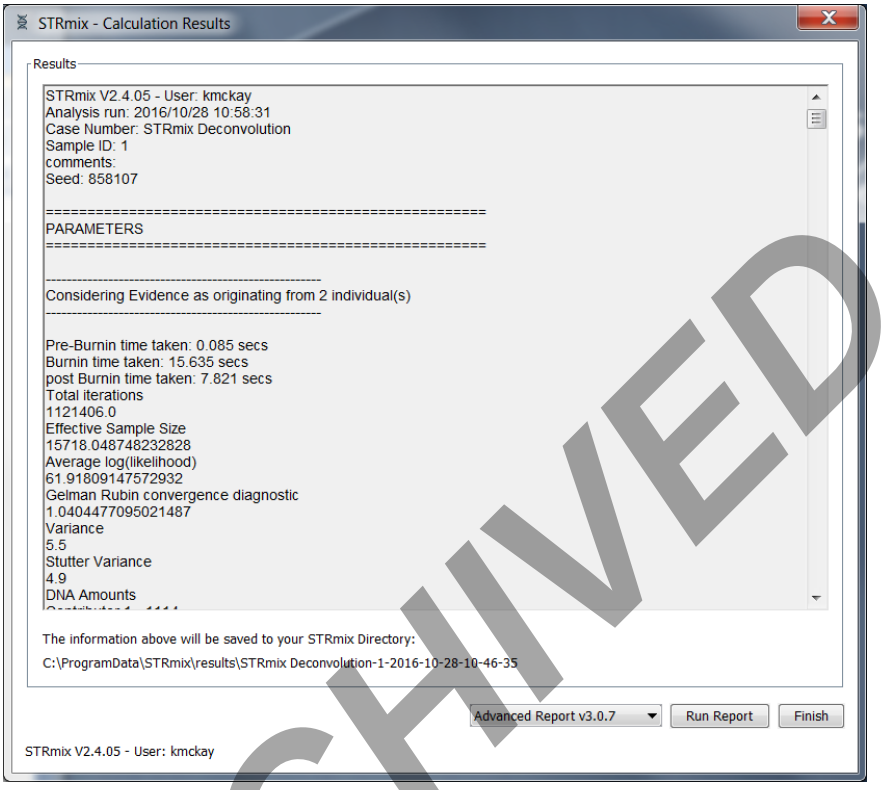
6. Select “Confirm Settings”. This will open up the Populations Settings window. When a reference sample is conditioned, population data is needed in the calculation. Ensure that the following four populations are listed: NIST Caucasian, NIST Asian, NIST AfAm, and NIST Hisp and that the settings appear as below.



H. After selecting "Start", the "Calculation Progress" window will open showing the Burnin progress and Main MCMC Progress.

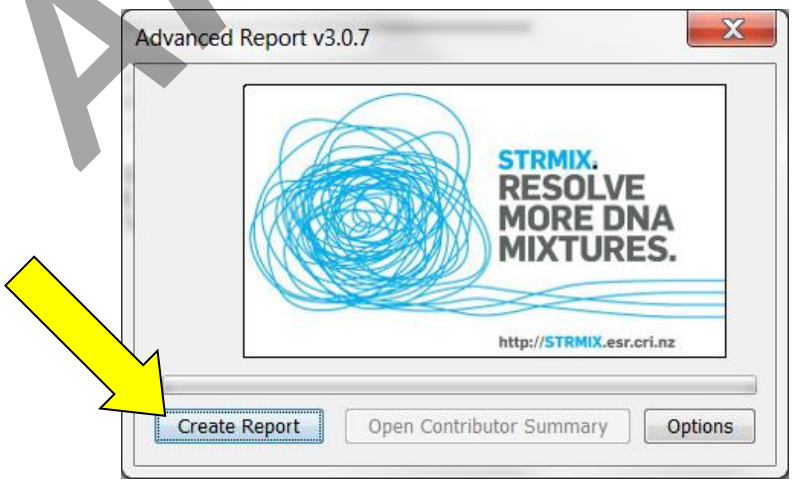


I. When the analysis is complete, the raw data report will open as follows:



J. Select "Run Report" from this screen to create the advanced report PDF which will be saved in the STRmix results folder for this analysis. Choose "Create Report" from the screen below

***CAUTION: If you hit "Finish" a Run Report will NOT be generated, and the analysis will have to be re-done.**



K. The analysis run name will auto-populate based on your Case and Sample ID naming. Hit “Save". The advanced report PDF will now open. Print the report.

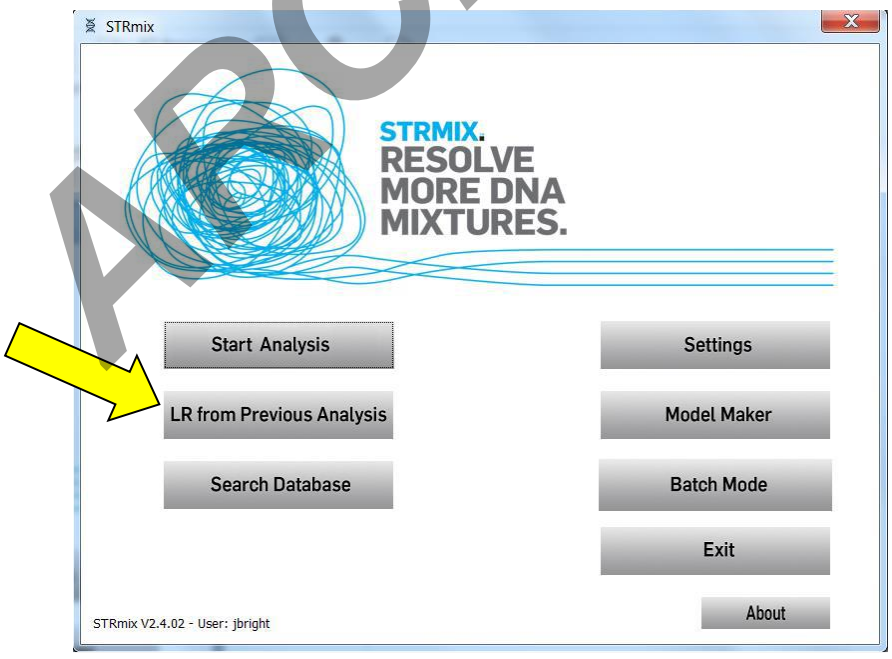
L. Close out of the Advanced Report window and then select "Finish" on the "STRmix - Calculation Results" window. This will return you to the STRmix start up screen.

M. Navigate to your STRmix Run Folder within the STRmix Results folder**. MOVE your STRmix Run folder** into the previously created FB sample folder within the STRmix Data folder.

II. Likelihood Ratio calculations with STRmix™:

Note: All samples must undergo deconvolution prior to running an LR with a comparison sample. Refer to the STR Results Interpretation Manual for guidance on when a conditioned contributor may be applied. A deconvolution of the evidence sample without conditioning followed by an LR against a potential conditioned contributor may need to be run first.

A. Launch the STRmix™ application. Open the STRmix software by locating STRmix in the task bar or by double clicking on the STRmix icon on the desktop. Select "LR from Previous Analysis".



B. Navigate to the folder where the STRmix deconvolution on the M drive (STRmix data) for the relevant sample is saved. Double click on the "settings.ini" file for the sample to select it.

C. The STRmix - "Configure Analysis" window will open.

D. Naming STRmix™ Likelihood Ratio runs

The Case Number and Sample ID will auto-populate from the deconvolution settings file. **This must be updated to the appropriate naming convention for an LR run before proceeding.** Use the following naming convention:

1. Evidence Files
   a. **Case Number** = YY-XXXXX (leave out "FB")
   b. **Sample ID** = remainder of the evidence sample OCME ID*_scenario for LR
   c. **Case notes** = leave blank

e.g. LR of 2 person mixture, FB16-01234_567_1_1.1_shirt_BL, comparing to elim John Doe
Case Number = 16-01234
Sample ID = 567_1_1.1_shirt_BL JD1Uv2U

2. Suspect Files
   a. **Case Number** = SYY-XXXXX (leave out "FB")
   b. **Sample ID** = OCME ID* for evidence sample (leave out "FB")_scenario for LR
   c. **Case Notes** = leave blank

*Suffixes such as 'mcon' or 'reamp' should not be included in the OCME ID

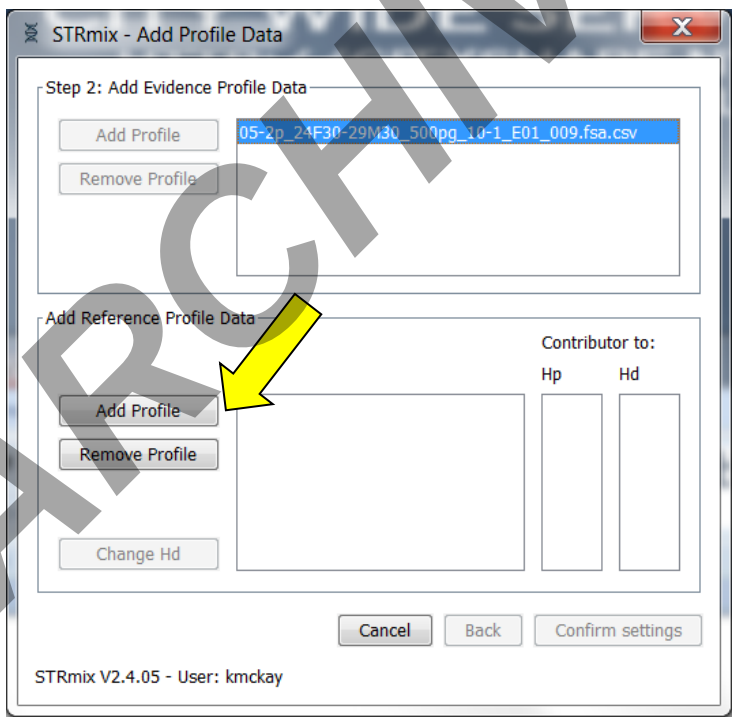
e.g. LR, FBS16-05678 suspect Tom Smith, comparing to 3 person mixture, FB16-01234_567_2_1.1_slide_GS
Case Number = S16-05678
Sample ID = 16-01234_567_2_1.1_slide_GS TS2Uv3U
Note – naming format for the LR scenarios should start with the comparison sample's initials, followed by any conditioned samples' initials, and then the number and "U" for unknowns, followed by a "v" to separate the numerator from the denominator hypotheses

Examples	**Resulting STRmix file name**
Evidence File	
elimAB vs 1 unknown	16-01234-567-1-1-1-shirt-BL-ABv1U
elimAB+2 unknowns vs 3unknowns	16-01234-567-1-1-1-shirt-BL-AB2Uv3U
elimAB+cond elim CD vs cond elim CD+1 unknown	16-01234-567-1-1-1-shirt-BL-ABCDvCD1U
Suspect File	
suspTS vs 1 unknown	S16-05678-16-01234-567-2-1-1-slide-GS-TSv1U

suspTS+1 unknown vs 2 unknowns	S16-05678-16-01234-567-2-1-1-slide-GS-TS1Uv2U
suspTS+cond elim CD vs cond elim CD+1 unknown	S16-05678-16-01234-567-2-1-1-slide-GS-TSCDvCD1U
suspTS+cond elim AB+1 unknown vs cond elim AB+2 unknowns	S16-05678-16-01234-567-2-1-1-slide-GS-TSAB1UvAB2U

E. Select "Confirm". **Once "Confirm" is hit, a folder will be created in the STRmix results folder on your C drive. If incorrect nomenclature was used and you return to this screen to make changes, the empty folder on your STRmix results drive should be deleted.**

F. In the "Add Profile Data" window, this is where you will import comparison input files and set hypotheses. Select "Add Profile" and navigate to the .txt file for the comparison samples and select the file(s). Alternatively, you may drag and drop .txt files of your reference samples into the Reference Profile Data box. Click "Add Profile Data".

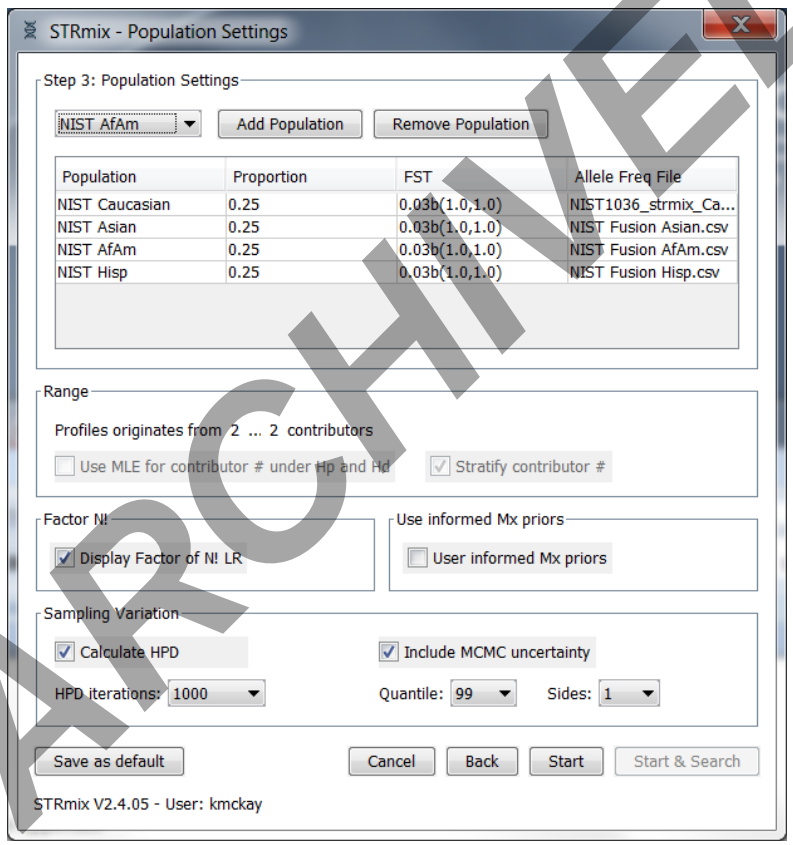


G. Each comparison sample (suspect, elim, or informative victim), that is added will be within the numerator of the LR (ie. assigned to H_p). If adding another comparison sample, repeat steps E and F for that comparison sample.

H. NOTE: If you are conditioning on a contributor, that conditioned sample should already be in the numerator (Hp) and denominator (Hd) from the deconvolution. Conditioning profiles may not be added at the LR step.

I. Then select "Confirm settings".

J. The Population Settings window will open. Ensure that the following four populations are present in the list: NIST Caucasian, NIST Asian, NIST AfAm, NIST Hisp, as seen below. Ensure that the settings are as pictured in the screen shot below.



K. Select "Start" to calculate the LR.

L. The Calculation Progress screen will open and the software will progress through burnin and Main MCMC Progress.

M. The STRmix calculations Results window will then open. Choose "Run Report" ***CAUTION: If you exit "Finish" a Run Report will NOT be generated, and the analysis will have to be re-done.**

N. The Advanced Report window will open. Select "Create Report".

O. The STRmix PDF report will open and will save in the relevant folder and then close the report.

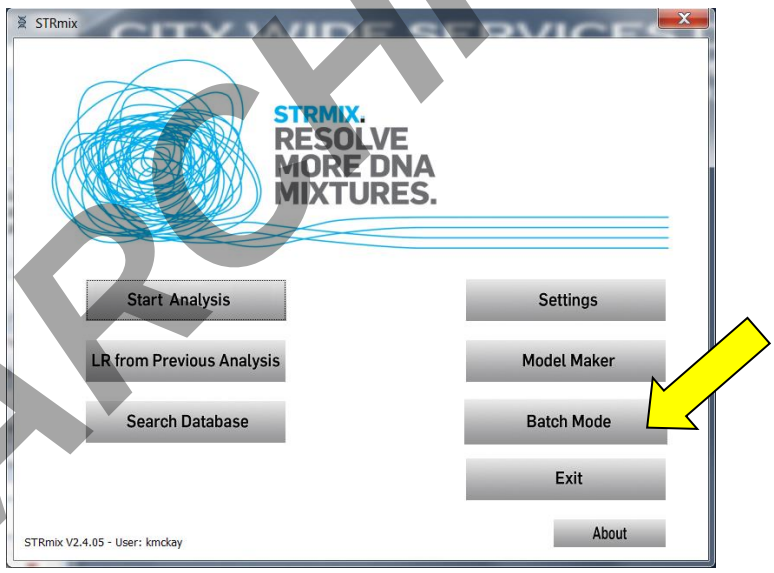
P. Close out of the Advanced Report window and then select "Finish" on the "STRmix - Calculation Results" window. This will return you to the STRmix start up screen.

Q. Navigate to your Run Folder in the STRmix Results folder on your C drive**. MOVE your STRmix Run folder** into the previously created FB Case folder in the STRmix Data folder.

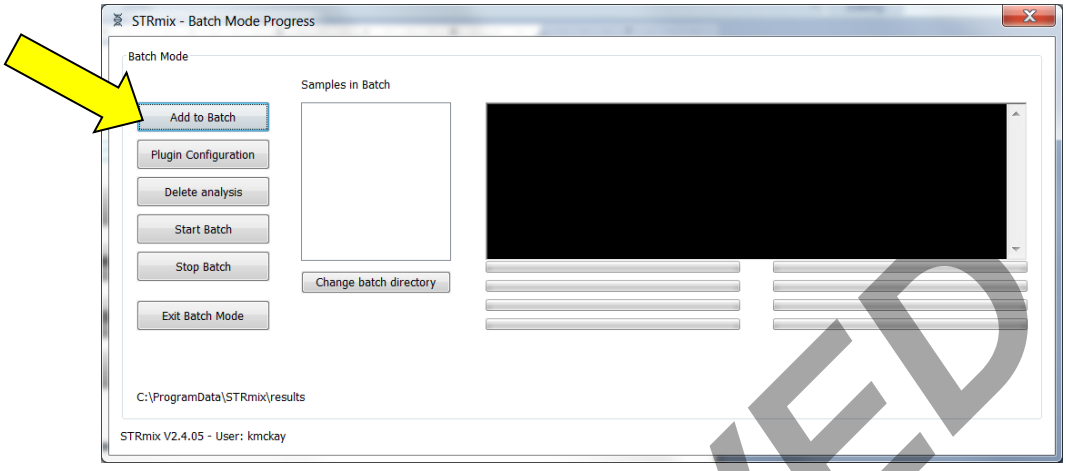
IV. How to Run STRmix™ using Batch Mode

A. A number of STRmix deconvolutions can be set up and queued to run sequentially. Note, if STRmix™ cannot model the data by the chosen NOC, the batch mode will stop and **no data will be saved**.

B. To set up a queued analysis for multiple samples, select Batch Mode from the STRmix main window.



C. Select “Add to Batch” from the Batch Mode window to open the Sample Summary window.



D. Complete the analysis set up for the first sample following steps **II. Deconvolutions in STRmix™**

E. In the Population Settings window, select "Start" to return to the Batch Mode window.

F. In the Batch Mode Window, select "Add to Batch" to enter the next sample. Repeat steps IV.C-E to add additional samples.

Note: to remove a sample from the batch mode, highlight the case/sample in the "Samples in Batch" section of the "Batch Mode" window then select "Delete analysis".

G. Select "Start Batch" to start the batch run.

H. After completion of analyses, select "Exit Batch Mode" to return to the STRmix™ main window.

I. Results folders from Batch Mode will be saved in the STRmix Results folder on your C drive. Move the results folders from the C drive to the appropriate FB folders within the STRmix Data drive.

V. Evaluation of the STRmix™ Analysis

A. Verify input files against the evidence input section of the STRmix™ printout and the electropherograms
   1. Correct input file(s) have been selected
   2. All appropriate edits were made, no artifact peaks were left labeled
   3. No allelic or stutter peaks have been removed

B. The number of contributors that best describes the data has been chosen

C. The correct assumptions (conditioning) have been made, if applicable

D. The appropriate proposition has been selected, if applicable

E. Check the Parameters table against the settings listed below to ensure that the correct settings were used for the run (note that any edited values are bolded by the program).

For Interpretation (Deconvolutions)

Setting	Value	Setting	Value	Setting	Value
Allele Variance	9.1374,0.7472 mode=6.080	Drop-in Cap	100.0	HPD Iterations	0
Stutter Variance	1.5007,12.9748 mode=6.496	Drop-in Frequency	0.0024	HPD Significance Value	0.0
Minimum allowed variance from the mode	0.5	Drop-in Parameters	0.0,0.0	HPD Sides	0
Loci	23	RWSD	0.005	Alleles Per Locus	15
Locus Amplification Variance	0.0065	ESS Thinning	100000	Factor of N!	Yes
Maximum	0.3	MCMC Accepts	500000	MCMC Uncertainty	Yes

Stutter					
Forward Stutter Max	0.1	Maximum Degradation	0.01	Burn-in Accepts	100000
Excluded Loci	DYS391	Saturation	8000	Chains	8

For Likelihood Ratios (Comparison)

Setting	**Value**	**Setting**	**Value**	**Setting**	**Value**
Allele Variance	9.1374,0.7472 mode=6.080	Drop-in Cap	100.0	HPD Iterations	1000
Stutter Variance	1.5007,12.9748 mode=6.496	Drop-in Frequency	0.0024	HPD Significance Value	0.99
Minimum allowed variance from the mode	0.5	Drop-in Parameters	0.0,0.0	HPD Sides	1
Loci	23	RWSD	0.005	Alleles Per Locus	15
Locus Amplification Variance	0.0065	ESS Thinning	100000	Factor of N!	Yes
Maximum Stutter	0.3	MCMC Accepts	500000	MCMC Uncertainty	Yes
Forward Stutter Max	0.1	Maximum Degradation	0.01	Burn-in Accepts	100000
Excluded Loci	DYS391	Saturation	8000	Chains	8

F. Verify that the following make intuitive sense when compared to the electropherogram(s):
   1. The mixture proportions assigned to the contributor(s)
   2. The weights assigned to the genotypes listed in the Genotype Probability Distribution
   3. The degradation values and Locus Efficiencies (LSAE)

G. Evaluate the following diagnostics for the run information. Note that the presence of a single sub-optimal diagnostic is not necessarily an indication that rework is required. In some instances sub-optimal diagnostics may be due to poor quality data and not due to an issue with the run.

   1. **Total number of iterations and acceptance rate.** A very low acceptance rate (e.g. 1 in thousands to millions) may, in combination with the other diagnostics, indicate that the analysis needs to be run for additional iterations.

   2. **Effective sample size (ESS).** A low ESS in relation to the total number of iterations suggests that the MCMC has not moved very far with each step or has a low acceptance rate. A low

ESS value (tens or hundreds) means that there is potential for a large difference in weights if the analysis is run again.

3. **Average (log) likelihood.** The larger this value, the better STRmix™ has been able to describe the observed data. A negative value suggests that STRmix™ has not been able to describe the data very well given the information it has been provided. A low or negative value for the average $\log_{10}$ (likelihood) may indicate to users that the analysis requires additional scrutiny.

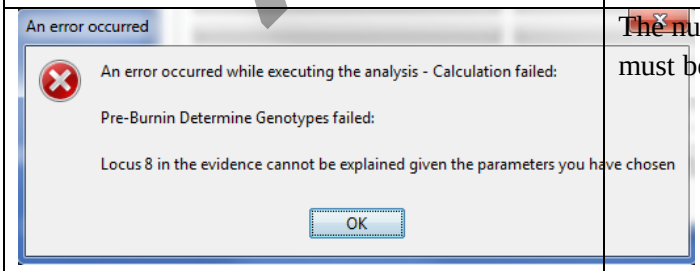
4. **Gelman-Rubin diagnostic.** If this value is above 1.2 then it is possible that the analysis has not converged.

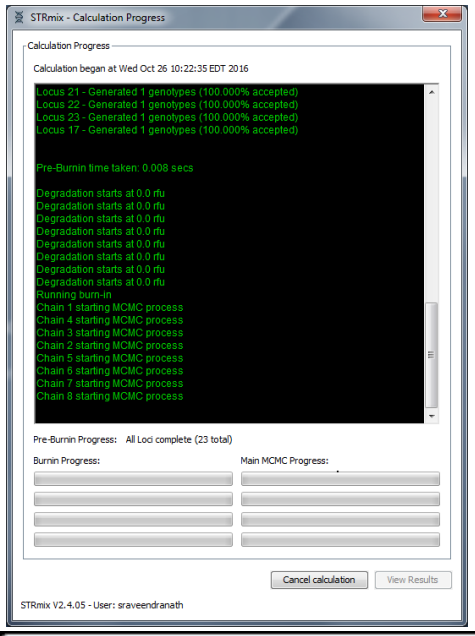
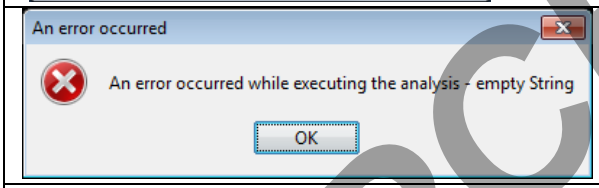
5. **Allele variance and stutter variance.** These variances should be compared to the mode. If the numbers are significantly elevated, the analysis may require additional scrutiny.

6. Evaluate the **Per Locus Likelihood Ratio** table per locus and per sample, if applicable; pay special attention to outliers and/or zero values.

Troubleshooting Guide

Observation	**Action**
A low or negative average (log) likelihood	Reevaluate number of contributors; consider another STRmix™ analysis with one higher or one lower number of contributors
	Data has been removed that is allelic and/or stutter, and must be re-imported
	Artifact peaks have been left labeled and must be removed
Gelman-Rubin value is greater than 1.2	Consider another STRmix™ analysis at greater number of iterations. **Note:** this requires approval by the Technical Leader
Stutter and/or allele variance significantly elevated above mode (may be in conjunction with low average (log) likelihood)	Check to make sure no data has been omitted
	Reevaluate number of contributors; consider another STRmix™ analysis with one higher or one lower number of contributors

	Consider amplifying a replicate if one has not already been done or increased input amount
Large LR's (>1) obtained for each locus, except one where the LR = 0 and the POI reference is consistent with the evidentiary profile	Data entry problem—check input files
	Allele call not fully resolved at a given locus – ignore locus and perform analysis again. **Note:** discuss with supervisor as needed
	Inhibition has occurred—microcon to clean and reamplify sample
	Consider amplifying a replicate if one has not already been done or increased input amount
	Reevaluate number of contributors; consider another STRmix™ analysis with one higher or one lower number of contributors
	Consider another STRmix™ analysis at greater number of iterations. **Note:** this requires approval by the Technical Leader
The mixture proportions do not reflect what is observed AND/OR the degradation does not reflect what is observed AND/OR the interpreted contributor genotypes do not make intuitive sense	Reevaluate number of contributors; consider another STRmix™ analysis with one higher or one lower number of contributors
	Inhibition has occurred—microcon to clean and reamplify sample
	Consider another STRmix™ analysis at greater number of iterations. **Note:** this requires approval by the Technical Leader
	The number of contributors selected was too few and must be increased
The STRmix™ run does not progress at the MCMC burn in stage and the chains do not	Check "alleles per locus" setting is set to 15.

move. Error message also received that Degradation started at "0". 	
	"Start & Search" was clicked instead of "Start". Analysis must set up and performed again.
Multiple diagnostics display sub-optimal results	Consider deeming the sample inconclusive, if all other actions have been exhausted

References:

STRmix™ v.2.4 Operation Manual

STRmix™ v. 2.4 Users Manual