DNA SOP-32 Using STRmix™ Software

32.1 Purpose:

STRmix™ is used to deconvolute an evidentiary profile to obtain probabilistic weights of genotype combinations for all contributors to either a single source or mixture samples with, or without conditioning known profiles. Those weights are used to calculate likelihood ratios which will be reported out on all appropriate samples deconvoluted through STRmix™. The association may be, based on the likelihood ratio, included, cannot be eliminated, inconclusive, or an elimination. STRmix™ may also be used as a tool to determine possible CODIS entries, though this is not currently validated for casework. For further guidance on running STRmix™ software, please refer to the current STRmix™ Operations Manual.

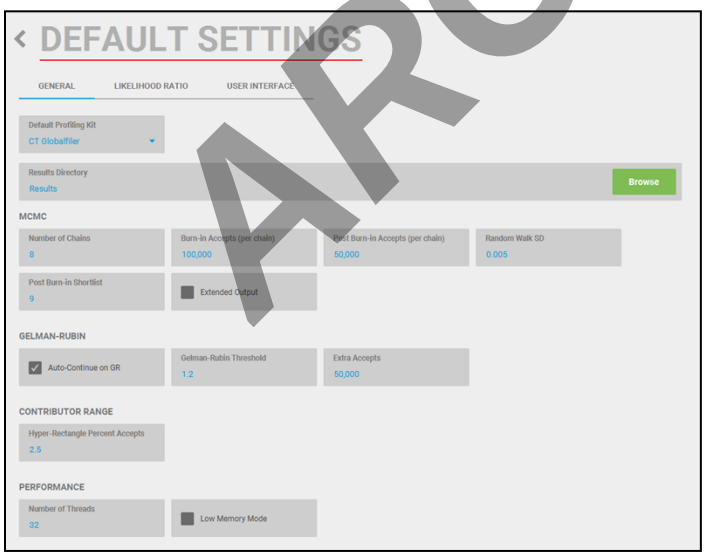
32.2 Responsibility:

DNA Unit personnel

32.3 Set-up of STRmix™

32.3.1 STRmix™ Default Settings

32.3.1.1 Default settings can be edited by clicking on "Administration" in the STRmix™ main menu, then clicking on "STRmix Defaults". See below for how your "Default Settings" should appear.



32.3.1.2 The number of MCMC chains, burn-in accepts, and post-burn-in accepts can be adjusted when scientifically valid (such as for a high Gelman-Rubin score or an expected genotype not being modeled) with TL approval.

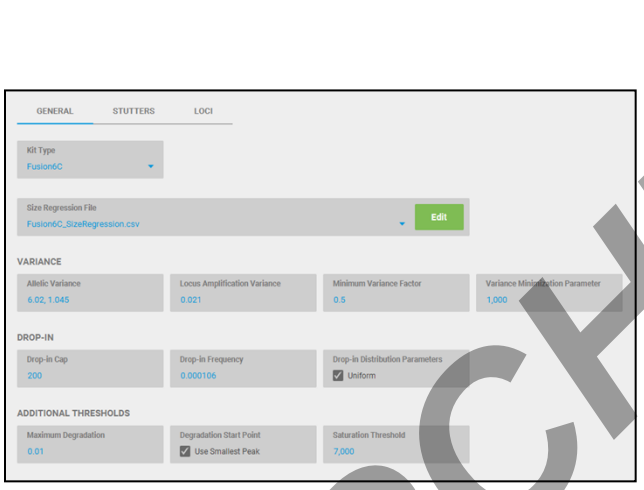
32.3.2 Kits

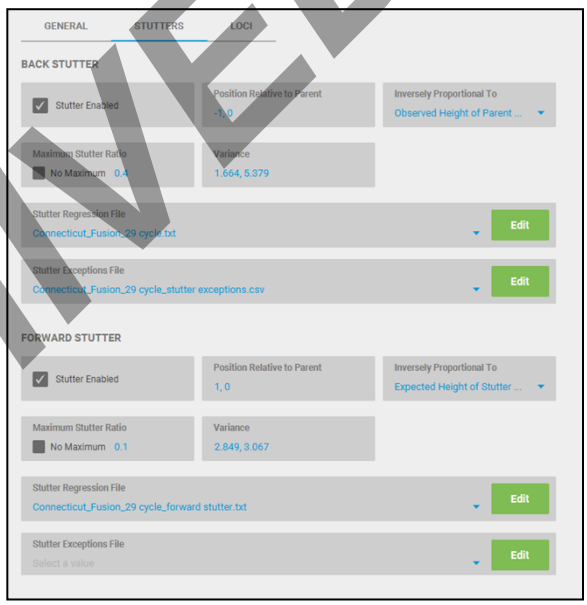
32.3.2.1 Kits can be edited by clicking on "Administration" in the STRmix™ main menu, then clicking on "Profiling Kits".

32.3.2.2 Located in server desktop's local disk ProgramData/STRmix/Kits folder will be 3 different kit files for Fusion 6C, for a maximum, standard, and low injection, 2 different kit files for Identifiler Plus (to be used for

Identifiler as well), for a maximum (15 seconds or more) and standard (less than 15 seconds) injection, and one kit file for GlobalFiler.

32.3.2.3 These are the settings for the F6C – Standard Injection kit, with the loci listed in order being AMEL (gender), D3S1358, D1S1656, D2S441, D10S1248, D13S317, Penta E, D16S539, D18S51, D2S1338, CSF1PO, Penta D, TH01, vWA, D21S11, D7S820, D5S818, TPOX, D8S1179, D12S391, D19S433, SE33, D22S1045, DYS391 (ignore), FGA, DYS576 (ignore), and DYS570 (ignore). Repeat length is 4 for all loci except D22S1045 (3), Penta E (5), and Penta D (5), with a detection threshold of 50 rfu for all loci, and back and forward stutter being modeled for all loci:





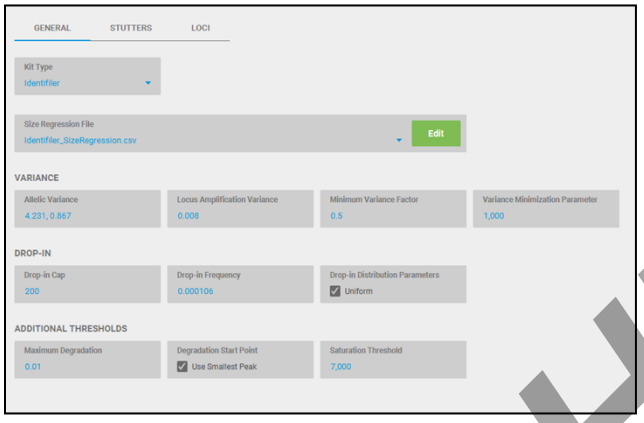
32.3.2.4 The F6C - Maximum Injection kit will have the same standard settings as the F6C – Standard Injection kit with the following exceptions:

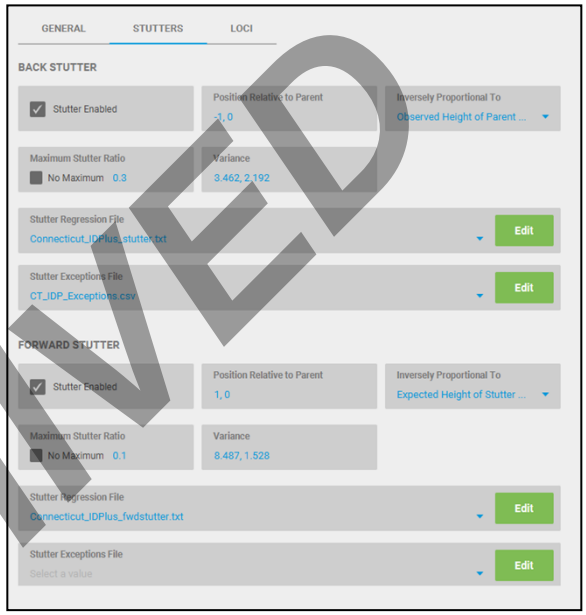
Allelic Variance: 6.725, 1.786
Stutter Variance: 2.666, 5.052
Locus Amp Variance: 0.026

32.3.2.5 The F6C – Low Injection kit will have the same standard settings as the F6C – Standard Injection kit with the following exceptions:

Allelic Variance: 4.129, 1.017
Stutter Variance: 1.75, 3.655
Locus Amp Variance: 0.019

32.3.2.6 These are the settings for the IDP – Standard Injection kit with the loci listed in order being D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, AMEL (gender), D5S818, and FGA. Repeat length is 4 for all loci, with a detection threshold of 50 rfu for all loci, and back and forward stutter being modeled for all loci:





32.3.2.7 The IDP – Max Injection kit will have the same settings as the IDP – Standard Injection with the following exceptions:

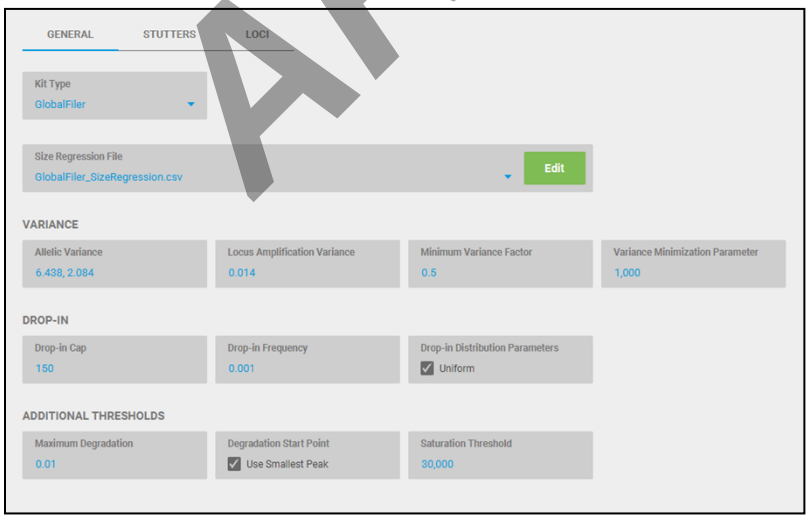
Allelic Variance: 6.262, 1.196
Stutter Variance: 4.218, 2.221
Locus Amp Variance: 0.01

32.3.2.8 These are the settings for the GlobalFiler injection kit (X is being modeled):

	Detection Threshold	Back Stutter	Forward Stutter	Half-Back Stutter	Double Back Stutter	Half-Forward Stutter
D3S1358	40	X	X	X	X	
vWA	40	X	X	X	X	
D16S539	40	X	X	X	X	
CSF1PO	40	X	X		X	
TPOX	40	X	X			
Yindel (Ignore)	55					

	Detection Threshold	Back Stutter	Forward Stutter	Half-Back Stutter	Double Back Stutter	Half-Forward Stutter
AMEL (Gender)						
D8S1179	55	X	X	X	X	
D21S11	55	X	X	X	X	
D18S51	55	X	X		X	
DYS391 (Ignore)	55					
D2S441	35	X	X	X	X	
D19S433	35	X	X	X	X	X
TH01	35	X	X	X		
FGA	35	X	X	X	X	X
D22S1045*	50	X	X		X	
D5S818	50	X	X		X	
D13S317	50	X	X		X	
D7S820	50	X	X		X	
SE33	50	X	X	X	X	X
D10S1248	55	X	X		X	
D1S1656	55	X	X	X	X	
D12S391	55	X	X		X	
D2S1338	55	X	X		X	

*Only 3 repeat locus. All others 4.



32.3.2.9 Certain situations (such as tri-alleles, discrepant results at one locus as compared to the rest of the profile, or the possibility for stutter exceeding a stutter max threshold) might warrant changing the "Ignore?" setting for autosomal loci during a deconvolution, with TL approval.

32.3.2.10 Please see the Appendix for screen shots of the stutter text files, and charts for the stutter exception files.

32.3.3 Populations/Allele Frequencies

32.3.3.1 Populations can be edited by clicking on "Administration" in the STRmix™ main menu, then clicking on "Populations" (see figure to right)

32.3.3.2 Each population file could only be assigned to one kit in STRmix™ v2.4, therefore, for backward compatibility with previous deconvolutions each F6C population file is in triplicate in the ProgramData/STRmix/Populations folder, and named to reflect the kit that it is assigned to. Population files for STRmix™ v2.5 and v2.6 are not kit-specific. Each file has an allele frequency file linked to it.

32.3.3.2.1 The allele frequency file that the African American population is linked to is FBI_extended_AfAmBahJam.csv

32.3.3.2.2 The allele frequency file that the Caucasian population is linked to is FBI_extended_Cauc.csf

32.3.3.2.3 The allele frequency file that the Hispanic population is linked to is FBI_extended_SE_Hisp.csv

32.3.3.3 Allele frequencies are from the FBI's allele frequencies in PopStats.

32.3.3.4 The allele frequencies listed in the CSV files do not have the posterior means equation applied to them.

32.3.3.5 Besides "Default FST" and the allele frequency file, most settings in "Populations" do not currently apply to State of Connecticut SOPs.

32.4 Initial Setup of Reports

32.4.1 Report Default Settings: From the Main Menu click "Administration", then "Report Defaults".

32.4.2 Only "Summary Report" should be clicked from the list on the left and set to contain the following options: Post Burn-In Summary, Contributor Summary, LR Summary, Component Information, Locus Efficiencies, Variance Charts, Settings, Interpretation Details, Probability of N Given the Profile, Evidence Files, and Reference Files. (The order of items in the "Include?" list can be changed by dragging.)

32.4.3 If viewing additional components for a run is necessary, reports can be re-created retroactively with different options selected. To do this click "Reports" from the main menu, and supply the appropriate run folder by browsing or drag-and-drop. Select the desired options, change the "Output File" name at the bottom of the window (as to not overwrite original report) and click "Run". This will not change the default report settings. Please note that the Genotype Probability Distribution is not included in the default report, so if there is a need to look at individual locus/contributor weights, a new report will need to be created. A report including all components shall be created for discovery/FOIA purposes, regardless if electronic data was requested.

32.4.4 With the settings above, the Interpretation Report is generated automatically for every deconvolution and LR from Previous run.

32.5 Using STRmix™ to deconvolute a profile

After analysis of a DNA profile, a determination of the most reasonable number of contributors is chosen and documented on DNA QR-302. Analysts manually compare all knowns associated with a case to all questioned samples associated with that case. If a positive association or inconclusive conclusion (except in cases of insufficient data or complexity of a mixture, where STRmix cannot be run) is made to a questioned DNA profile that questioned sample shall be deconvoluted using STRmix™. Identical profiles (regardless of injection time) compared to the same known, as such will give identical conclusions, need only be run once through STRmix. This will be properly documented in the DNA Report.

32.5.1 Launch STRmix™ software via remote desktop server: Whenever prompted, usernames, passwords, and settings can be stored.

32.5.1.1 Open the following website using your web browser: (Can be saved to favorites) https://10.51.107.36/RDWeb/Pages/en-US/login.aspx. This initially may come up as an unsecure website, but continue to the website when prompted by your browser. This link will bring up your remote application and desktop. If prompted to log in, Domain\username shall be entered as: dps\employee#. Password shall be the same password used to log onto your DPS desktop.

32.5.1.2 Click on "explorer". This will bring up the server desktop, or a prompt to save/open an application. Click "ok", "continue" or "confirm" whenever prompted. Two drives will be accessed through this desktop:

32.5.1.2.1 The F-drive is the STRmix™ server. It is the results folder on this drive to which your results will be automatically saved. Located in F:\results are individual analyst folders. Any deconvolutions an analyst performs are to be saved temporarily in these individual analyst folders. If these are not automatically mapped to save in that location, the analyst shall move

their results folders into their folder. Analysts further will create case specific folders in their results folders. This can be done by right-clicking in their folder → new → folder. The folder names will be DSS-YY-XXXXXX, for consistency and ease in archiving. Do not create any further sub-folders, as file name sizes are restricted. Upon issuance of a DNA report, this folder shall be moved to the F:\Results\ Completed folder. A record of this move shall be recorded on QR-4, and confirmed by your technical reviewer. There shall be no sub-folders in "Completed" to assist with archiving data.

32.5.1.2.2 The U: Drive that you see on the server desktop is the U: Drive that your GeneMarker/ GeneMapper text files have been previously exported and saved to. You will not be saving anything to the U: Drive while using STRmix™. However, you must access those files from the server desktop, or you will not be allowed by STRmix™ to drag and drop those text files.

32.5.1.3 Click on "STRmix". This will launch the STRmix™ software or a prompt to save/open an application. Click "ok", "continue" or "confirm" whenever prompted. Only 12 analysts can concurrently run the STRmix™ software.

32.5.2 Click "Interpretation" from the start menu.

32.5.3 Enter the following information:

Case number: Lab ID# (ex. DSS-XX-XXXX)

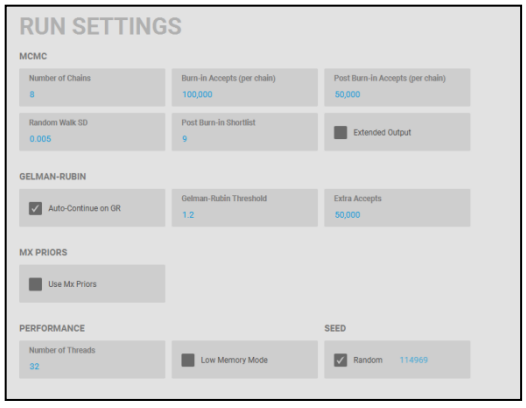
Sample ID: Item #(s) (ex. 1-1 LR to 2)

Case Notes: Add any information that would be important to your technical reviewer. This can include, but not be limited to description of the item, knowns, conditioning information, reason (if applicable) deconvolution is being repeated, etc.

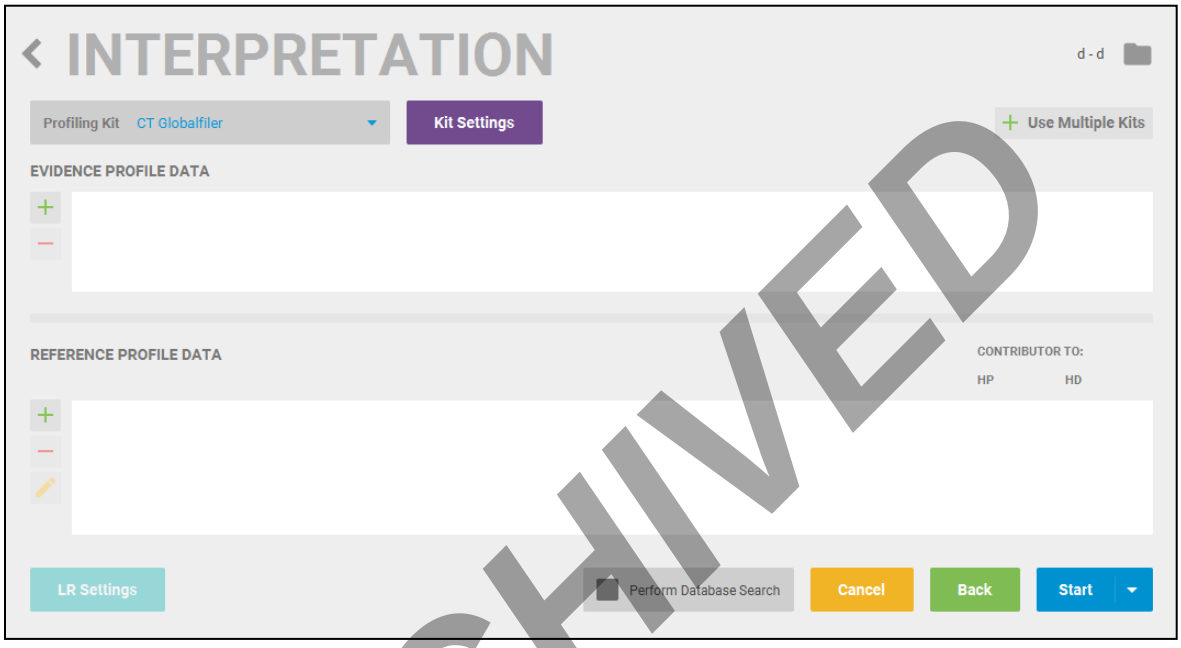
Number of contributors: This should have been pre-determined, and recorded on QR-302 during analysis. ("Contributor Range" should never be selected, as this feature is not currently validated CT DSS.)

32.5.3.1 A click on "Run Settings" will bring up the window to the right, Nothing in here need be changed in routine casework.

32.5.3.2 After everything has been properly added, click "Next". (Note: At any time you can return to earlier steps using the "Back" button. All entries will be saved.)



32.5.4 In window shown below, first select the kit appropriate to the amplification system and injection time for the evidentiary sample from the "Profiling Kit" pull-down menu. The "CT GlobalFiler" kit will be there as a default.



32.5.4.1 Drag text file with evidentiary file results (the text file exported from GeneMarker/ GeneMapper) to be deconvoluted into "Evidence Profile Data" box (or conversely select green plus sign and browse to file). If necessary, browse through the list of samples in that file, select the one you are deconvoluting by checking the box to the left of it, and click "Confirm". The evidentiary input file HAS stutter peaks (these stutter peaks HAVE NOT been filtered out).

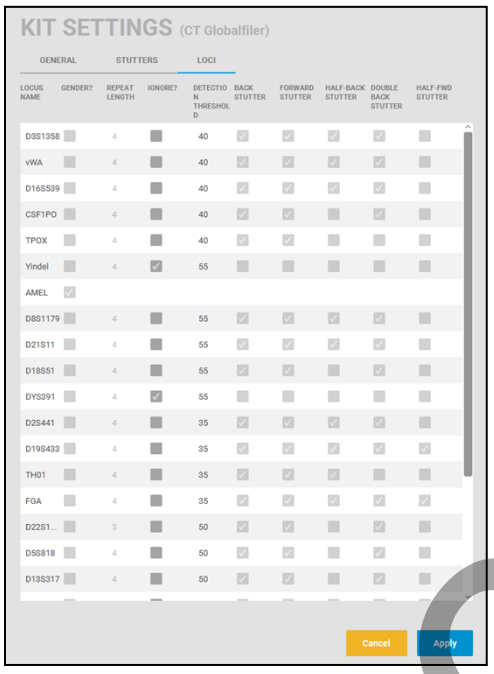
32.5.4.2 Add reference profiles to the "Reference Profile Data" box in the same way as above. The known input file does NOT have stutter peaks (all stutter peaks have been filtered out).

32.5.4.3 For any .csv input files that were generated for STRmix v2.4, STRmix v2.6 will prompt you for the kit used to type the sample being added. Providing any STRmix kit name that uses sample's amplification system will suffice for this; there is no need to look up the injection time used. This step is needed because these input files indicate marker only by reference numbers based on the order of loci in the kit, rather than containing locus names. The kit selected in step 32.5.4 will still be the kit used for the deconvolution.

32.5.4.4 If you are conditioning (a) known(s) to H_d, click on the H_d box that corresponds to that known. See workflow at the end of SOP.

32.5.5 Ignoring Loci (with TL approval)

32.5.5.1 Click the "Kit Settings" button next to the profiling kit pull-down menu, then click on "Loci", to show the window to the left.



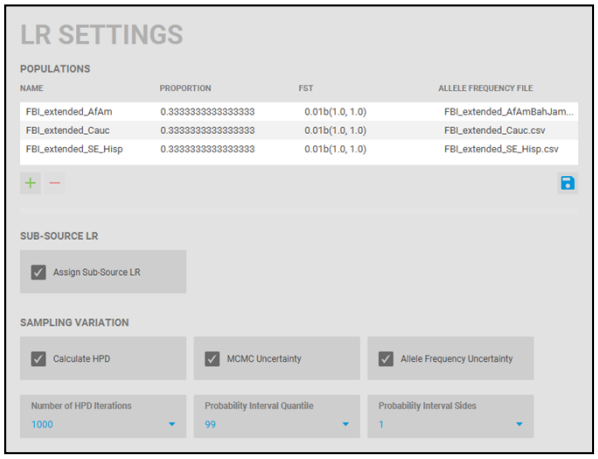
32.5.5.2 Check the box under "Ignore" for the locus you wish to ignore. Click "Apply" to change for this run only. Default settings will not be changed.

32.5.5.3 The stutter max thresholds are expected to be conservative for all loci in all kits.

32.5.6 To see the defaulted populations, click on "LR settings". The window below will appear. When compared knowns are applied to any hypothesis, 3 populations are added to "Population Settings": African American, Caucasian, and Hispanic. This setting is completed and saved the first time STRmix™ is run, and need not be completed on subsequent deconvolutions. Click "Start", making sure that "Perform Database Search" is not checked off.

32.5.7 The report PDF file for the deconvolution opens automatically when the run completes. Print page 1 for a deconvolution-only run, and print pages 1-3 for a run with a likelihood ratio calculation, for the case jacket.

32.5.8 The qualitative comparisons for various kits are as follows:



Profiling Kit	**Exclusion**	**Inconclusive**	**Positive Association**
Identifiler/Identifiler Plus	<1	≥ 1 <1,000	≥1,000
Fusion 6C	<1	≥ 1 <10,000	≥10,000
GlobalFiler	<1	≥ 1 <1,000	≥1,000

32.6 Using STRmix™ to Create a Likelihood Ratio to a Reference Sample from previously deconvoluted sample

32.6.1 After launching STRmix™, click on "Investigation" then "LR from Previous". Into the "Choose Previous Interpretation for LR" box, you can either drag and drop the outer run file folder of the deconvolution that was previously completed, or browse to select the "config" .xml file within that folder, then click "select".

32.6.2 Next the sample ID may be edited to reflect the current calculation. (i.e. 1-1 LR 2 changes to 1-1 LR 3). This can be further described in the case notes. Since this is to a previous deconvolution, the MCMC settings are unable to be changed. Click "Next".

32.6.3 You will be unable to add or remove evidence profiles, or change kits in analysis. Add reference profiles to the bottom half of the "add profile data" window, as in 32.5.4.2. The known can only be added to H_p. Knowns cannot be added to or removed from H_d for conditioning purposes in "LR from previous analysis" settings, since conditioning a known affects the deconvolution, and not just the likelihood ratio from previously obtained weights.

32.6.3.1 If rework on this sample is being completed because a victim, consensual partner, or other known has been submitted that is to be conditioned to the profile, the deconvolution itself must be repeated with the conditioned known.

32.6.4 Print pages 1-3 of each report for the case jacket.

32.7 Running samples in Batch Mode

32.7.1 On the STRmix™ startup window, click "Batch Mode".

32.7.2 Click the "Add to Batch" pull-down menu, and select either "Add Interpretation" or "Add LR From Previous", depending on the desired calculation, and add samples as you would in 32.5 or 32.6. After a sample has been added to the batch, you will not be able to see your settings. Pay close attention to your settings. Entering something incorrectly could stop the batch from running.

32.7.3 If a sample needs to be run in low memory mode (a complex mixture, for example), please adjust in "Run Settings" for that particular sample.

32.7.4 If all samples have been added, but you do not wish to start the batch at this time, you can click the back arrow next to "Batch Mode" to return to the main menu. The samples that have been added will be there when you return to "Batch Mode".

32.7.5 If you wish to start the batch, click "Start".

32.7.6 If you wish to end the batch, click "Stop". All samples that have yet to be deconvoluted will remain in the queue, until removed.

32.7.7 To remove a sample from the queue, highlight the sample in the "Calculations In Batch" window, then click "Remove".

32.8 **Review of Run Diagnostics:** In STRmix™, there are a number of diagnostics that may indicate that MCMC analysis did not perform as expected. These are found in the Post Burn-In Summary section of the Report.

For every deconvolution performed in STRmix, DNA-QR-303 must be completed to allow for an overall assessment of how the deconvolution was performed in STRmix. The ideal ranges for the quality parameters based on the internal validation are documented on this QR. No single diagnostic can determine whether a STRmix deconvolution was or was not successful. However, multiple values outside of their ideal ranges, or values falling far from their ideal ranges, are potential indicators of a problem with a deconvolution, and may warrant manual scrutiny of the genotypic weights assigned and possibly a re-run with TL approval.

In the event that any of the secondary diagnostic values are out of range, the analyst will document a brief justification in the "Comments" section of QR-303 for why the deconvolution is (or is not) valid.

32.8.1 Total iterations

The value indicates the total number of post burn-in iterations that the MCMC ran for during its analysis. This value, along with the number of accepts chosen for the analysis informs the analyst as to how often a new proposed set of parameters was accepted. This is referred to as the acceptance rate. For example, the analysis that lead to the above results was carried out with 100,000 burn-in accepts and 500,000 total accepts. This leaves 400,000 post burn-in accepts spread across 800,000 total iterations giving an acceptance rate of 1 in 2.

A very low acceptance rate (e.g. 1 in hundreds or thousands) may, in combination with the other diagnostics, indicate that the analysis needs to be run for additional iterations. On its own, a low acceptance rate is not an indication that reanalysis is required.

As a note, if a deconvolution auto-continues past the standard number of post burn-in accepts (due to an above-threshold Gelman-Rubin diagnostic score), STRmix v2.5.11 factors the total number of post burn-in iterations but only the standard number of accepts into the acceptance rate on the STRmix Report. The macro in QR-303 uses the total post burn-in counts for both iterations and accepts when calculating acceptance rate. The value appearing on the worksheet is the appropriate one to compare to the threshold range. This has been corrected in STRmix v 2.6.2.

32.8.2 Effective Sample Size

Effective sample size (ESS) is the number of independent genotypic/mass parameter combinations that the MCMC has evaluated. 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 had a low acceptance rate. A low ESS (e.g. 10s or 100s) value means that there is potential for a large difference in weights if the analysis was run again. A low ESS on its own is not an indication that reanalysis is required. In general, an ESS of less than 750 warrants additional scrutiny.

32.8.3 Average ($\log_{10}$) likelihood

This value is the average ($\log_{10}$) likelihood for the entire post burn-in MCMC. It is the log of the average likelihood (or probability) value created at each of the post burn-in MCMC iterations. 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. Some possible reasons for this value being low or negative are:

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

b) The number of contributors is wrong which can cause STRmix™ to consider incorrect genotypic combinations (e.g. large heterozygote peak imbalances or variation in mixture proportions across the profile).

c) Data has been removed that was real, particularly stutter peaks, and must now be described in STRmix™ by dropout.

d) Artifact peaks were not removed and must now be accounted for in STRmix™ by drop-in.

A low or negative value for the average ($\log_{10}$) likelihood may indicate that the analysis requires additional scrutiny.

Good quality (adequate template, high molecular weight) mixed DNA profiles often give higher average ($\log_{10}$) likelihood values than comparable single source profiles. So low average ($\log_{10}$) likelihood values alone are not necessarily an indicator of an issue especially if the profile is single source.

32.8.4 Gelman-Rubin convergence diagnostic

This diagnostic informs the analyst whether the MCMC analysis has likely converged. STRmix™ uses 8 chains to carry out the MCMC analysis and ideally each chain will be sampling in the same space after burn-in. If the chains spend their time in different spaces then it is likely that the analysis has not run for long enough. Whether or not the chains have spent time in the same space can be gauged by the within-chain and between-chain variances. This is known as the Gelman-Rubin convergence diagnostic (GR). If the chains fully converge, the GR is 1.

If the GR is above 1.2 (or 1.35 for 4 or 5 person mixtures), then there exists the possibility that the analysis hasn't converged. If the GR value is above 1.2 (or above 1.35 for 4 or 5 person mixtures), the results of the analysis should be scrutinised. Running the analysis for a larger number of iterations may reduce the GR in these instances. With the default settings a run will continue automatically for an additional 50,000 accepts per chain if the GR score is above 1.2 after the standard accept count; this does not preclude analysts from increasing the iteration count further if appropriate.

32.8.5 Allele Variance and Stutter Variance constants

Both of these values are the average value for variance and stutter variance constants across the entire post burn-in MCMC analysis. These values can be used as a guide as to the level of stochastic variation in peak heights that is present in the profile. A visual representation of how these values compare with the data seen in the laboratory's model maker data can be seen in "Variance Charts" section in the STRmix Report. The dot on the curve represents the value in the current run.

If variance is significantly above the mode value, additional scrutiny is warranted. It may indicate that the DNA profile is sub-optimal or that the number of contributors is incorrect. A higher than average Stutter Variance constant may be indicative of stutter peaks not being called in the STRmix GeneMarker project. A further look into the STRmix GeneMarker project may be necessary to see if all stutter peaks have been properly called. Pay close attention to those peaks adjacent to another peak with only a 1 base pair difference, since those are the more likely to be unresolved than those with greater than a 1 base pair difference.

Used in conjunction with the average ($\log_{10}$) likelihood, a large allele variance or stutter variance may indicate that the PCR did not perform as expected.

If the sample is simply low level, this often results in a low average ($\log_{10}$) likelihood and an average variance constant.

If some data has been omitted or mistakenly left in the STRmix™ input file, this often results in a low average ($\log_{10}$) likelihood and high variances.

32.9 Using STRmix to perform a staff search

To search the staff index in STRmix, set up the run the same way as for an interpretation (see 32.5) but check the box "Perform Database Search" before clicking start. If the interpretation has already be done, choose "Database Search" from the "Investigation" menu and drag the run folder into the "Previous Interpretation" window. The staff search database settings will auto-populate by default. The database search report will return any matches to the staff index with a LR ≥ 1000 (without HPD).

Glossary of Terms

Accept: An iteration that is accepted. Default settings in STRmix™ require 500,000 accepts in the MCMC process. The burn-in, the first 100,000 accepts, allows the MCMC chains to proceed into spaces where "good" genotypes are proposed. The Metropolis-Hastings algorithm determines if a new genotype combination will be accepted or rejected.

Allele Variance: A secondary diagnostic tool, that when compared to the kit's mode, can determine how variable peak heights are in a particular sample, as well as to those samples most commonly seen in the laboratory's model maker.

Average (log) likelihood: A secondary diagnostic tool that calculates the average of the common logarithm of all likelihoods obtained for the post burn-in MCMC process.

Effective Sample Size (ESS) – A secondary diagnostic tool that evaluates the total number of independent genotypic/mass parameter combinations that the MCMC has evaluated.

Gelman-Rubin convergence diagnostic: A secondary diagnostic tool that compares variances within and between MCMC chains, to determine if they have or haven't converged.

Iteration: a proposed genotype combination.

Likelihood Ratio: A statistical test for comparing two competing hypotheses.

Markov chain Monte Carlo (MCMC): An algorithm based on standard mathematical principles that assigns a likelihood for each random genotype combination, in order to provide the best explanation for an observed data set.

Mass Parameters: 4 variable parameters used in the MCMC process to generate expected peak heights given a proposed genotype combination. These are: contributor specific **D**egradation rate, locus specific **A**mplification efficiencies, PCR **R**eplicate (not utilized by CT DSS SOPs), and contributor specific **T**emplate amount. A mnemonic phase to easily remember these parameters is **DART**.

Model Maker: A tool within the software that uses laboratory specific empirical data to create kit specific STRmix™ parameters.

Probabilistic Genotyping: A tool that combines the use of biological modeling, statistical theory, computer algorithms, and probability distributions to calculate likelihood ratios for DNA typing results of forensic samples, providing statistical weighting to different genotype combinations

STRmix™: Forensic analysis software that utilizes a fully continuous (taking peak heights into account) approach to probabilistic genotyping to deconvolute DNA profiles into their most probable components,

then compare reference samples to that deconvolution, calculating a likelihood ratio to give that comparison a statistical weight.

Stutter Variance: A secondary diagnostic tool, that when compared to the kit's mode, can determine how variable stutter peak heights are in a particular sample, as well as to those samples most commonly seen in the laboratory's model maker.

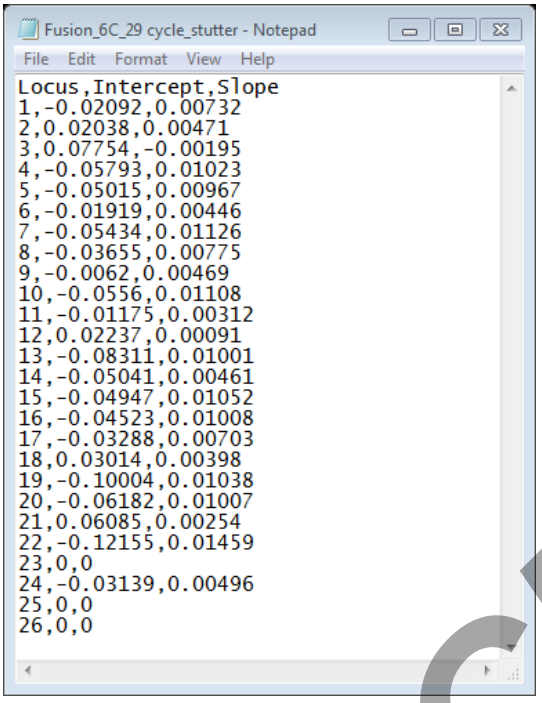
Total Iterations: A secondary diagnostic tool that tells how many genotype combinations were proposed during the post burn-in MCMC process. Directly related to the acceptance rate, which divides the total number of post burn-in accepts (defaulted to 400,000) by the total number of iterations. Therefore, if the total iterations were 7 million, the acceptance rate would be 1 in 17.5.

Weight: A probability given to a particular genotype combination that correlates to how well that combination explains the observed profile.

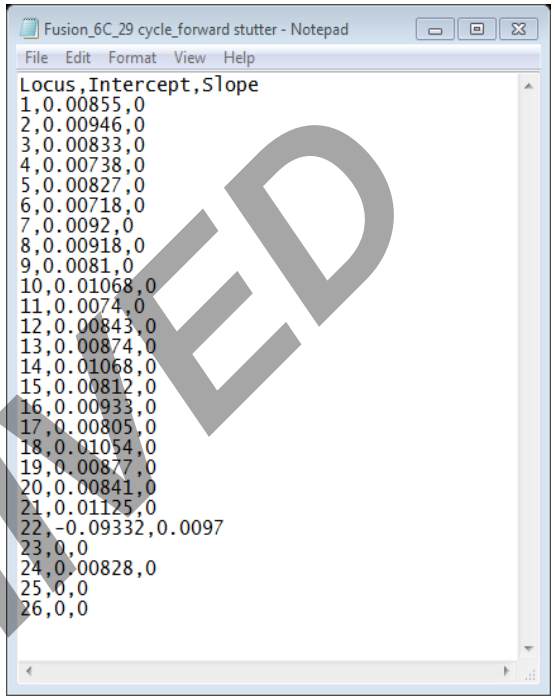
Note: In creating the above glossary, the following references were utilized: STRmix Training Workshop handouts, NYC OCME STRmix™ Glossary, the STRmix Support Website, the STRmix Operations Manual, and the STRmix™ User's Manual.

Appendices

Appendix 1: F6C Stutter File



Appendix 2: F6C Forward Stutter File



Appendix 3: F6C Stutter Exceptions File:

What is shown in this appendix is a truncated version of the stutter exceptions file, listing only all of the stutter exceptions. The full version of this file lists all loci and all repeats in those loci.

Repeat	D3S1358
13	0.06355
14	0.07844
15	0.09291
16	0.0933
17	0.10154
18	0.11101
19	0.1177

Repeat	D1S1656
11	0.06336
12	0.0734
13	0.08337
14	0.09366
15	0.10519
15.3	0.06924
16	0.11508
16.3	0.07731
17	0.12828
17.3	0.08733
18.3	0.0956
19.3	0.11022

Repeat	D2S441
10	0.04354
11	0.06581
11.3	0.0307
12	0.07349
14	0.04604
15	0.05485

Repeat	D2S1338
16	0.07082
17	0.07211
18	0.07885
19	0.08786
20	0.09115
21	0.0908
22	0.08697
23	0.09679
24	0.10597
25	0.12193
26	0.11848

Repeat	CSF1PO
7	0.02889
8	0.0446
9	0.03888
10	0.05278
11	0.06379
12	0.08068
13	0.09312

Repeat	Penta D
9	0.02391
10	0.02139
11	0.02226
12	0.023
13	0.02947
14	0.03312
15	0.03318

Repeat	TH01
6	0.02126
7	0.03261
8	0.03284
9	0.03904
9.3	0.01725

Repeat	D21S11
27	0.06409
28	0.077798
29	0.083498
30	0.094359
30.2	0.082745
31	0.10219
31.2	0.086892
32	0.110109
32.2	0.096907
33.2	0.10599
35	0.088913

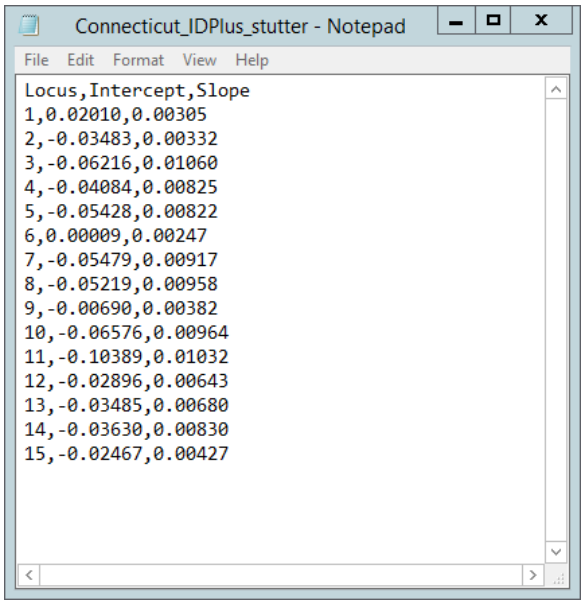
Repeat	D8S1179
9	0.05481
10	0.06995
11	0.08066
12	0.08394
13	0.07828
14	0.08691
15	0.0869
16	0.09698

Repeat	D19S433
11	0.04463
12	0.0616
13	0.06884
13.2	0.07345
14	0.07851
14.2	0.09034
15	0.08811
15.2	0.09411
16	0.09051
16.2	0.10265

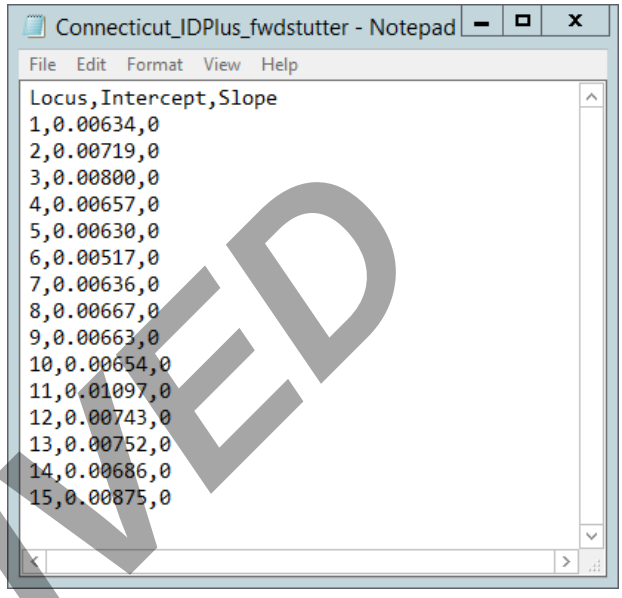
Repeat	SE33
13	0.07679
14	0.08063
15	0.08763
16	0.10257
17	0.10101
18	0.11568
19	0.12278
20	0.12455
21	0.1356
21.2	0.08847
22	0.13586
22.2	0.09782
23.2	0.10263
24.2	0.1108
25.2	0.10887
26.2	0.12206
27.2	0.12681
28.2	0.13212
29.2	0.13372
30.2	0.14036
31.2	0.15647
32.2	0.15267

Repeat	FGA
19	0.05522
20	0.0641
21	0.0703
22	0.08002
23	0.08644
24	0.09284
25	0.09859
26	0.10413
27	0.09188

Appendix 4: ID/IDP Stutter File



Appendix 5: ID/IDP Forward Stutter File



Appendix 6: ID/IDP Stutter Exceptions File

What is shown in this appendix is a truncated version of the stutter exceptions file, listing only all of the stutter exceptions. The full version of this file lists all loci and all repeats in those loci.

Allele	D21S11
27	0.047908
28	0.054311
29	0.062966
30	0.070931
30.2	0.052559
31	0.077371
31.2	0.062083
32	0.08521
32.2	0.071276
33.2	0.080434
35	0.045518

Allele	TH01
4	0.00245
5	0.00759
6	0.01273
7	0.01787
8	0.02301
8.3	0.00759
9	0.02815
9.3	0.01273
10	0.03329
10.3	0.01273
11	0.03843
12	0.04357
13.3	0.02301

Allele	D2S1338
16	0.049954
17	0.057781
18	0.065838
19	0.06834
20	0.07168
21	0.074861
22	0.070758
23	0.074588
24	0.085853
25	0.091494
26	0.09872

Allele	D19S433
6.2	0.00402
8	0.01308
9	0.01308
10	0.0312
11	0.04026
11.1	0.0312
11.2	0.04932
12	0.04932
12.1	0.00402
12.2	0.05838
13	0.05838
13.2	0.06744
14	0.06744
14.2	0.0765
15	0.0765
15.2	0.08556
16	0.08556
16.2	0.09462
17	0.09462
17.2	0.09915
18	0.10368
18.2	0.11274
19.2	0.1218

Allele	FGA
18.2	0.050704
19	0.046956
20	0.052723
21	0.06027
22	0.074733
23	0.077721
24	0.082077
25	0.09072
26	0.090288
27	0.079836

Appendix 7: **GlobalFiler Stutter Regression Files (0,2), (1,0), (-1,0), (-1,2) & (-2,0)**

ConnecticutGF3500_(0,2)_Regression

File Edit Format View Help

```
Locus,Intercept,Slope
D3S1358,0,0
vWA,0,0
D16S539,0,0
CSF1PO,0,0
TPOX,0,0
Yindel,0,0
D8S1179,0,0
D21S11,0,0
D18S51,0,0
DYS391,0,0
D2S441,0,0
D19S433,0.00793,0
TH01,0,0
FGA,0.00749,0
D22S1045,0,0
D5S818,0,0
D13S317,0,0
D7S820,0,0
SE33,0.00795,0
D10S1248,0,0
D1S1656,0,0
D12S391,0,0
D2S1338,0,0
```

ConnecticutGF3500_(1,0)_Regression

File Edit Format View Help

```
Locus,Intercept,Slope
D3S1358,0.00777,0
vWA,0.00621,0
D16S539,0.00807,0
CSF1PO,0.00787,0
TPOX,0.00425,0
Yindel,0,0
D8S1179,0.00829,0
D21S11,0.00984,0
D18S51,0.00870,0
DYS391,0,0
D2S441,0.00799,0
D19S433,0.00703,0
TH01,0.00341,0
FGA,0.00839,0
D22S1045,-0.05928,0.00651
D5S818,0.01011,0
D13S317,0.00871,0
D7S820,0.00677,0
SE33,0.01102,0
D10S1248,0.00735,0
D1S1656,0.00995,0
D12S391,0.00916,0
D2S1338,0.00987,0
```

ConnecticutGF3500_(-1,0)_Regression

File Edit Format View Help

```
Locus,Intercept,Slope
D3S1358,-0.04101,0.00725
vWA,-0.07068,0.00816
D16S539,-0.04508,0.00872
CSF1PO,-0.04372,0.00881
TPOX,-0.02447,0.00526
Yindel,0,0
D8S1179,0.01232,0.00370
D21S11,-0.05012,0.00396
D18S51,-0.03297,0.00655
DYS391,0,0
D2S441,0.05263,-0.00064
D19S433,-0.06695,0.00959
TH01,0.01153,0.00106
FGA,-0.06418,0.00599
D22S1045,-0.11235,0.01267
D5S818,-0.03519,0.00804
D13S317,-0.05057,0.00872
D7S820,-0.04024,0.00816
SE33,0.03781,0.00236
D10S1248,-0.02509,0.00689
D1S1656,0.02003,0.00343
D12S391,-0.08615,0.00846
D2S1338,-0.01485,0.00439
```

ConnecticutGF3500_(-1,2)_Regression

File Edit Format View Help

```
Locus,Intercept,Slope
D3S1358,0.00405,0
vWA,0.00449,0
D16S539,0.00453,0
CSF1PO,0,0
TPOX,0,0
Yindel,0,0
D8S1179,0.01331,0
D21S11,0.00537,0
D18S51,0,0
DYS391,0,0
D2S441,0.00757,0
D19S433,0.00671,0
TH01,0.00405,0
FGA,0,0
D22S1045,0,0
D5S818,0,0
D13S317,0,0
D7S820,0,0
SE33,0.04165,0
D10S1248,0,0
D1S1656,0.01731,0
D12S391,0,0
D2S1338,0,0
```

ConnecticutGF3500_(-2,0)_Regression

File Edit Format View Help

```
Locus,Intercept,Slope
D3S1358,0.00517,0
vWA,0.00593,0
D16S539,0.00403,0
CSF1PO,0.00321,0
TPOX,0,0
Yindel,0,0
D8S1179,0.00526,0
D21S11,0.00719,0
D18S51,0.00637,0.00073
DYS391,0,0
D2S441,0.00355,0
D19S433,0.00519,0
TH01,0,0
FGA,0.00553,0
D22S1045,0.00760,0
D5S818,0.00565,0
D13S317,0.00516,0
D7S820,0.00566,0
SE33,0.00717,0
D10S1248,0.00620,0
D1S1656,0.00651,0
D12S391,0.00752,0
D2S1338,0.00653,0
```

Appendix 8: **GlobalFiler Stutter Exceptions File**

What is shown in this appendix is a truncated version of the stutter exceptions file, listing only all of the stutter exceptions. The full version of this file lists all loci and all repeats in those loci.

Allele	D3S1358
14	0.057157213
15	0.070214002
16	0.075747885
17	0.083579096
18	0.086703111
19	0.09515721

Allele	vWA
11	0.021647446
14	0.039794758
15	0.05419452
16	0.061073697
17	0.067741678
18	0.076002233
19	0.084110913
20	0.094287847

Allele	D16S539
8	0.027102144
9	0.033088544
10	0.043033654
11	0.04957558
12	0.060711435
13	0.068121727
14	0.075502413

Allele	D8S1179
8	0.030275779
9	0.041166635
10	0.048186229
11	0.05877945
12	0.056821559
13	0.060381419
14	0.064776565
15	0.066719985
16	0.068078219
17	0.076700175

Allele	D21S11
26	0.045255334
27	0.050348048
28	0.058834134
29	0.066059096
30	0.075774293
30.2	0.059425255
31	0.079463683
31.2	0.069985026
32	0.090087629
32.2	0.077065249
33.2	0.083942629
35	0.067970436

Allele	D2S441
10	0.035079766
11	0.054400091
11.3	0.024948462
12	0.059030209
13	0.035869284
14	0.041421602
15	0.04730129

Allele	D19S433
6.2	0.00497
8	0.0136
9	0.0136
10	0.03086
11	0.03949
11.1	0.03086
11.2	0.04812
12	0.04812
12.1	0.00497
12.2	0.05675
13	0.05675
13.2	0.06538
14	0.06538
14.2	0.07401
15	0.07401
15.2	0.08264
16	0.08264
16.2	0.09127
17	0.09127
17.2	0.095585
18	0.0999
18.2	0.10853
19.2	0.11716

Allele	D5S818
8	0.02578807
9	0.035359618
10	0.047376719
11	0.053247853
12	0.061253483
13	0.067554112
14	0.079974379

Allele	TH01
4	0.00356
5	0.00903
6	0.0145
7	0.01997
8	0.02544
8.3	0.00903
9	0.03091
9.3	0.0145
10	0.03638
10.3	0.0145
11	0.04185
12	0.04732
13.3	0.02544

Allele	FGA
18	0.047377555
19	0.045966818
20	0.052071688
21	0.0627042
21.2	0.065717445
22	0.067640009
23	0.07592171
24	0.08225654
25	0.087313067
26	0.092385336
27	0.073225804

Allele	D13S317
8	0.018639161
9	0.027636656
10	0.036289016
11	0.046703807
12	0.053198383
13	0.064382035
14	0.068873716

Allele	D1S1656
10	0.038091039
11	0.052122684
12	0.058509489
13	0.066715776
14	0.076017969
14.3	0.044753523
15	0.079769254
15.3	0.055455096
16	0.096188462
16.3	0.061246782
17	0.105950187
17.3	0.068641764
18.3	0.07755221
19.3	0.084337418
20.3	0.097746847

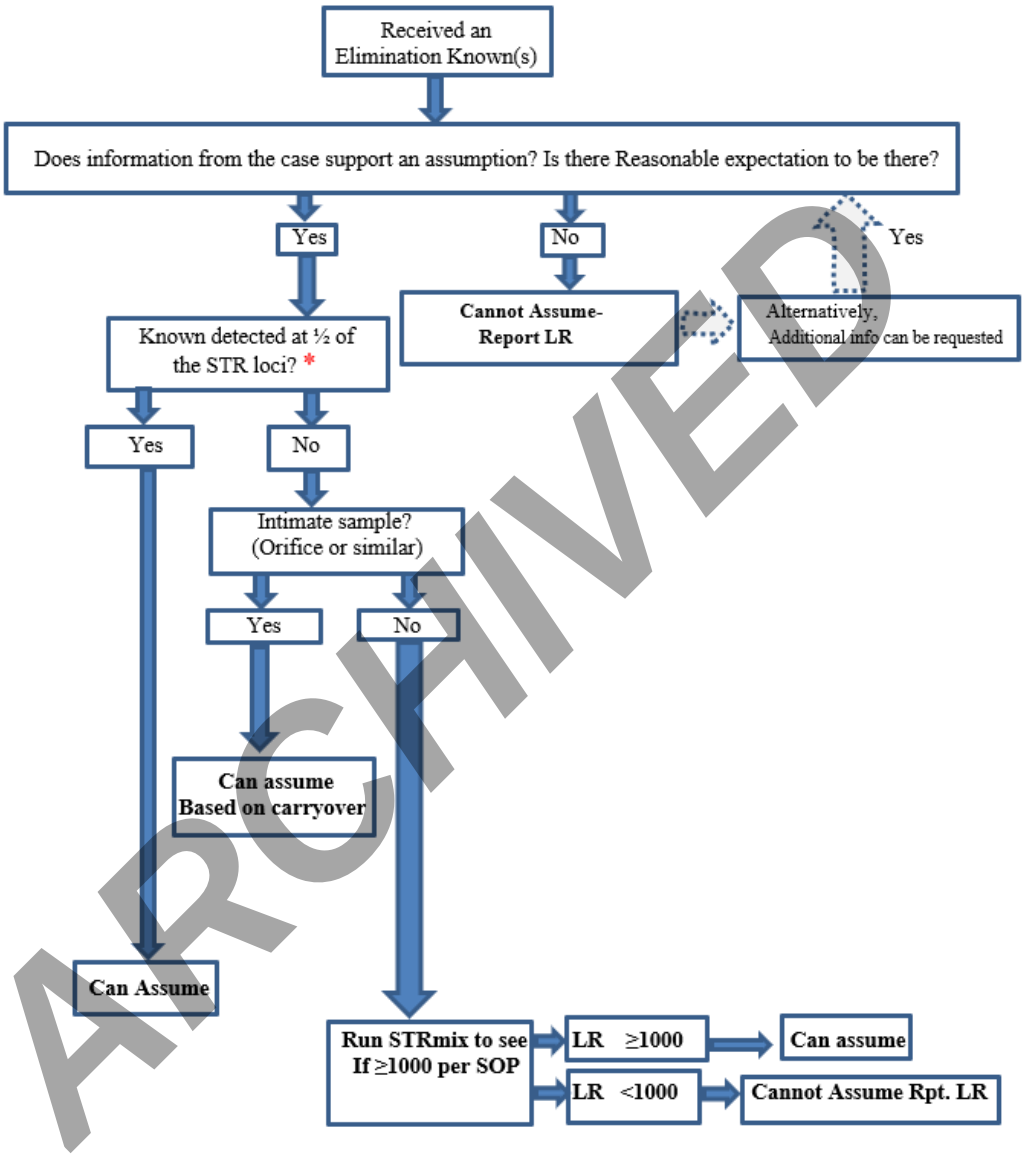
Allele	D10S1248
11	0.045478762
12	0.057250355
13	0.065590485
14	0.070354463
15	0.082042273
16	0.081661818
17	0.096379288

Allele	SE33
12	0.051066981
13	0.057925719
14	0.060510302
15	0.069474101
16	0.074128367
17	0.081578331
18	0.089001609
19	0.089724768
19.2	0.057838049
20	0.099225077
20.2	0.065143047
21	0.112380789
21.2	0.060111285
22	0.113095663
22.2	0.0695048
23.2	0.074836329
24.2	0.081108059
25.2	0.085765998
26.2	0.089076674
27.2	0.099475408
28.2	0.103651021
29.2	0.111906187
30.2	0.114894038
31.2	0.121249021
32	0.100129656
32.2	0.12989792
33	0.090853457
33.2	0.131442883
34	0.096257412

Allele	D12S391
15	0.039509506
16	0.051744278
17	0.058346473
17.3	0.052592996
18	0.068585161
18.3	0.060679173
19	0.07982839
20	0.082668327
21	0.088835583
22	0.098274453
23	0.108756947
24	0.119927917
25	0.130966451

Allele	D2S1338
16	0.050174213
17	0.058245042
18	0.063006326
19	0.072160424
20	0.080432849
21	0.075459958
22	0.07820601
23	0.078338562
24	0.090973177
25	0.093499216
26	0.108471302
27	0.110911362

Workflow for Elimination Knowns vs. Assuming/Conditioning.



***If you have a 4–5-person mixture: needs to be more loci and definitive due to coincidental matches. Use caution. See TL for any questions.**