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 we will report out on all 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. For further guidance on running STRmix™ software, please refer to the STRmix™ v 2.4 Operations Manual dated 6 July 2016.

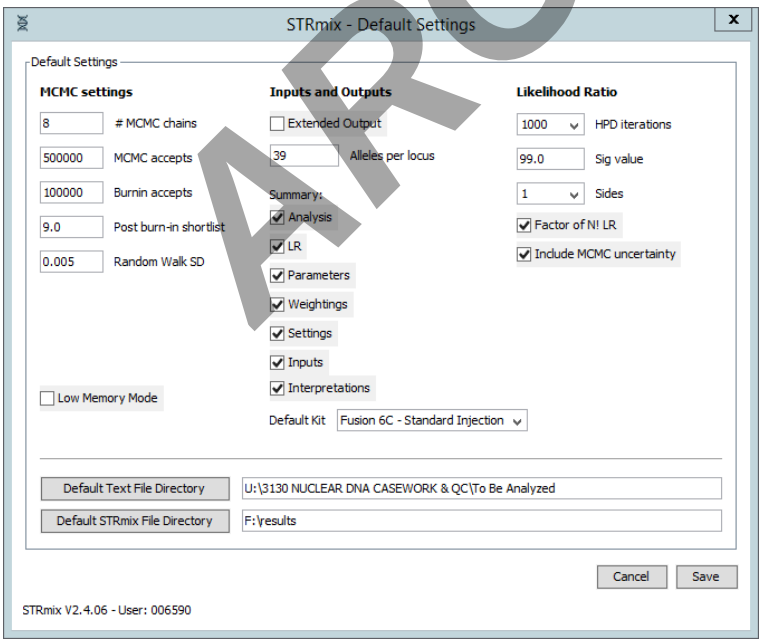
32.2 Responsibility:

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 "Settings" in the STRmix™ main menu, then clicking on "Set STRmix™ Defaults". See below for how your "Default Settings" should appear.



32.3.1.2 The number of MCMC chains, MCMC accepts, and Burn-in accepts can be adjusted when scientifically valid and with TL approval.

32.3.2 Kits

32.3.2.1 Kits can be edited by clicking on "Settings" in the STRmix™ main menu, then clicking on "Add/Edit DNA profiling kit"

32.3.2.2 Located in server desktop's local disk (C:) ProgramData/STRmix™/Kits folder will be 3 different kit files for Fusion 6C: Maximum Injection, Standard Injection, and Low injection. Each 3130 will have three allowed injection times. Please see SOP 30.5. 2.4 for these injection times. Note that the utilization of the maximum injection is with TL or designee approval.

- The Maximum Injection kit will be used whenever amplification product is injected on a 3130 at the maximum allowable injection time on that instrument.
- The Standard Injection kit will be used whenever amplification product is injected on a 3130 at the standard injection time on that instrument.
- The Low Injection kit will be used whenever amplification product is injected on a 3130 at the low injection time on that instrument.

32.3.2.3 All three kits will have the following, standard settings. These will not change, with the exceptions of that asterisked. Please see the Appendix for screen shots of the stutter text files.

- Stutter File: Fusion_6C_29 cycle_stutter.txt
- Stutter Exceptions File: Fusion_6C_29 cycle_stutter exceptions.csv
- Forward Stutter File: Fusion_6C_29 cycle_forward stutter.txt
- Number of loci: 27
- Gender Locus: AMEL
- Locus Order: AMEL,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,FGA,DYS576,DYS570
- *Include Loci: Y,Y,Y,Y,Y,Y,Y,Y,Y,Y,Y,Y,Y,Y,Y,Y,Y,Y,Y,Y,Y,N,Y,N,N
- Stutter Threshold: 50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50
- Stutter Max: 0.4
- Forward Stutter Max: 0.1
- Degradation starts at: -1.0
- Degradation Max: 1
- Drop-in cap: 200
- Drop-in parameters: 0,0
- Saturation: 7000
- Var > mode: 0.5
- Drop-in Frequency: 1.06E-4

* Certain situations might warrant the changing of this setting, based on expertise, extenuating circumstances, and with TL approval

32.3.2.4 The maximum injection kit will have the following standard settings. These will not change.

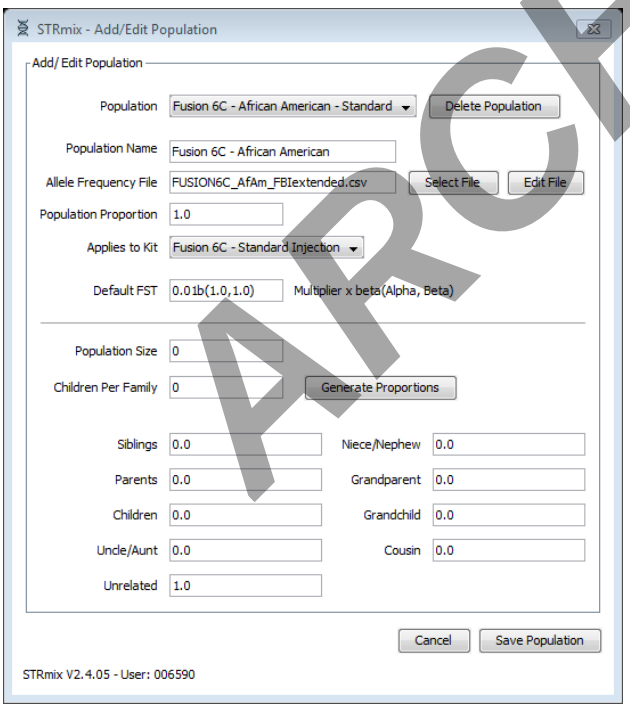
- Allelic Variance: 8.83,1.38
- Stutter Variance: 6.35,1.96
- Locus Amp Variance: 0.027

32.3.2.5 The standard injection kit will have the following standard settings. These will not change.

- Allelic Variance: 6.73,1.01
- Stutter Variance: 2.97,3.06
- Locus Amp Variance: 0.02

32.3.2.6 The low injection kit will have the following standard settings. These will not change.

- Allelic Variance: 3.20,1.37
- Stutter Variance: 2.83,2.77
- Locus Amp Variance: 0.021



32.3.3 <u>Populations/Allele Frequencies</u>

32.3.3.1 Populations can be edited by clicking on "Settings" in the STRmix™ main menu, then clicking on "Edit/View population" (see figure to left)

32.3.3.2 Each population file can only be assigned to one kit in STRmix™, therefore, each population file is in triplicate in the ProgramData/ STRmix/Populations folder, and named to reflect the kit that it will be assigned to. These files are assigned to the kits in the "Add/Edit Population" window. Each file has an allele frequency file linked to it.

32.3.3.2.1 The allele frequency file that the African American populations is linked to is: FUSION6C_AfAM_FBIextended.csv

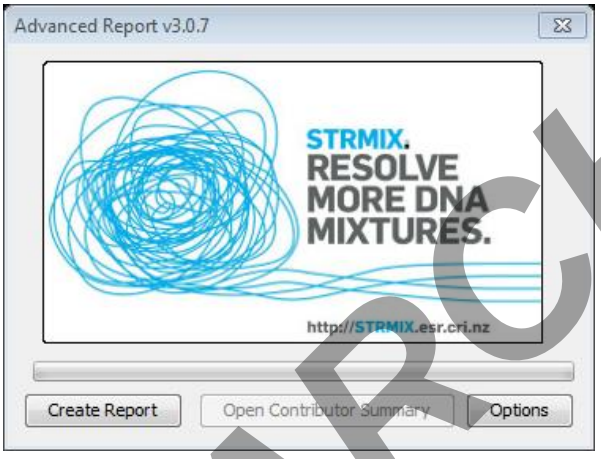
32.3.3.2.2 The allele frequency file that the Caucasian populations is linked to is: FUSION6C_CAUC_FBIextended.csv

32.3.3.2.3 The allele frequency file that the Hispanic populations is linked to is: FUSION6C_SEHISP_FBIextended.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.

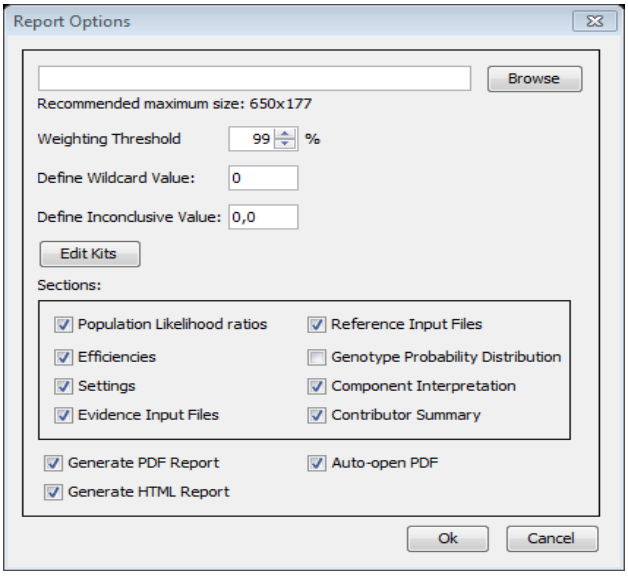
32.3.3.5 When looking at the "Add\edit Population" window, for each population group, the Population Name, Allele Frequency File, Population Proportion, and Default FST are the same regardless of kit. The bottom half of the screen does not currently apply to State of Connecticut SOPs.



32.4 Initial Setup of Advanced Reports

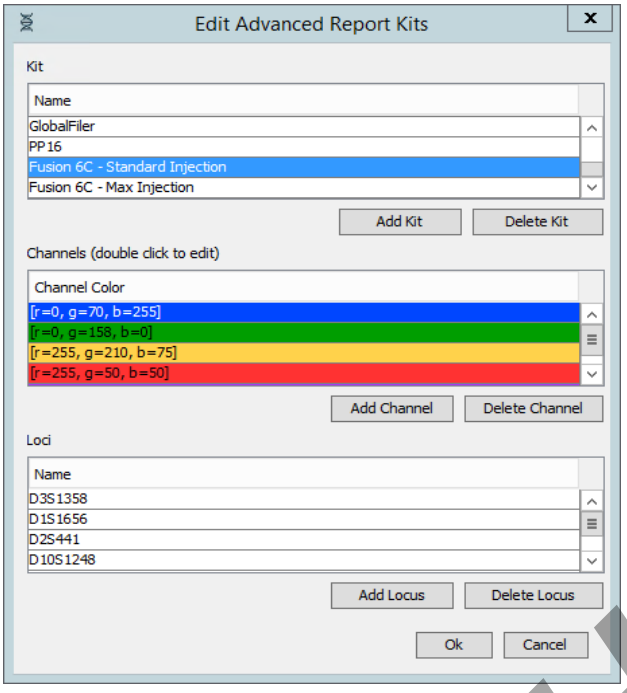
32.4.1 Setting Default Settings: After running a deconvolution, when the "Calculation Results" window opens, click on "Run Report". The "Advanced Report v3.0.7" window, shown to the left, then pops up.

32.4.2 To set up Advanced Reports initially, click on "Options". This need not be done for routine casework. Default settings are saved for future advanced reports.



32.4.3 The options as shown to the right have been selected as the laboratory's default Advanced Report settings.

32.4.4 Click on "Edit Kits" to bring up the window shown on the next page, to set up Locus Efficiencies table. Kit names must be identical to those in STRmix™. Channel color should correspond to color of dye in amplification kit, and loci names must be identical to those in the STRmix™ kits.



32.5 Using STRmix™ to deconvolute a profile

After analysis of a DNA profile, a determination of the most reasonable number of contributors (refer to SOP 31) 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 is made to a questioned DNA profile, whether it be single source or a mixture profile, that questioned sample shall be deconvoluted using STRmix™. Identical profiles compared to the same known, that will give identical conclusions, need only be run once through STRmix. This will be properly documented in the case jacket and 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.

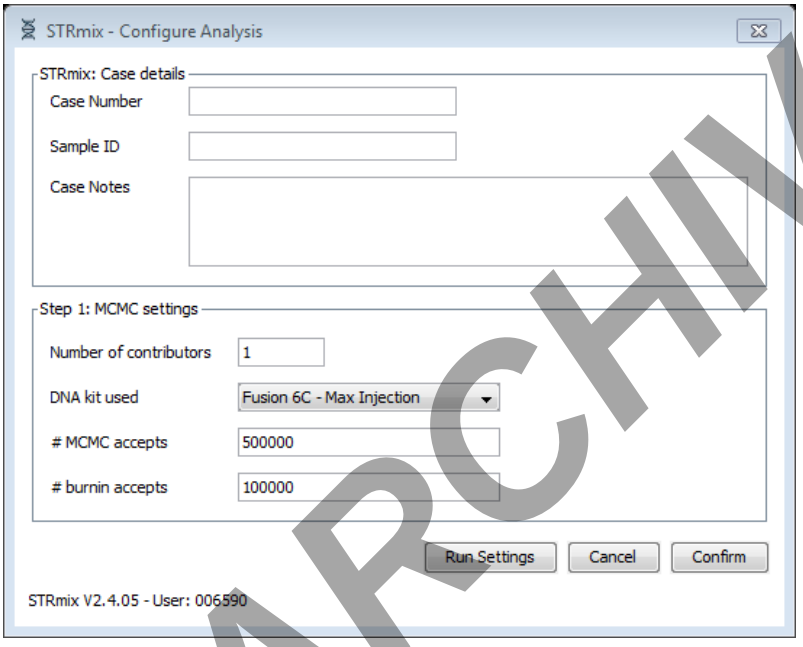
32.5.1.2 Click on "explorer". This will bring up the server desktop, or a prompt to save/open an application. Click "ok", "continue" or "confirm" whenever prompted. Two drives will be accessed through this desktop:

32.5.1.2.1 The F-drive is the STRmix™ server. It is the results folder on this drive where your results will be automatically saved to. 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. 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 backing up files/archiving data.

32.5.1.2.2 The U-drive that you see on the server desktop is the U-drive that your GeneMarker 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 "Start Analysis". The window to the left will appear.

32.5.3 Enter the following information into "Configure Analysis" Window:

Case number: Lab ID# (DSS-XX-XXXXXX)

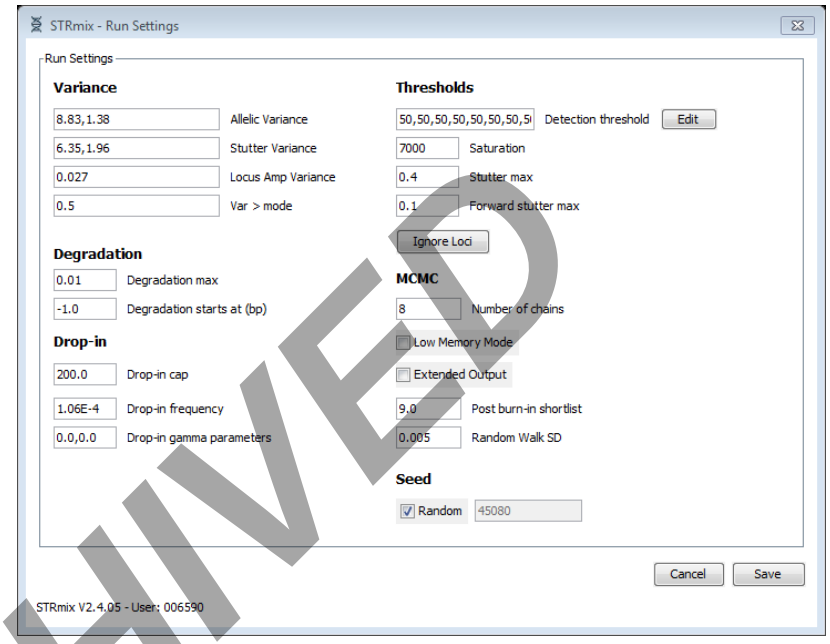
Sample ID: Item #(s) (1G1 LR to 2)

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.

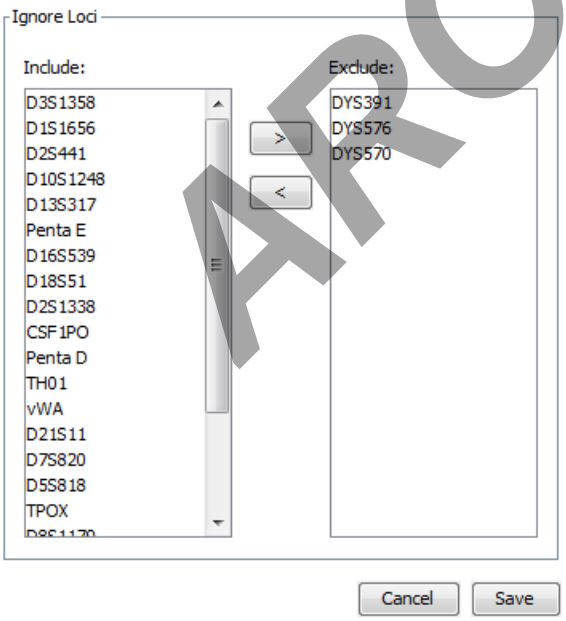
DNA Kit used: low, standard, or maximum injection kit, depending on the injection time of the sample.

# of accepts (MCMC and burn-in): default settings, need not be changed in routine casework. If warranted, and with TL approval, # of accepts can be changed here.



32.5.4 A click on "run settings" will bring up the window to the right: (variances may be different depending on chosen kit)

Nothing in here need be changed in routine casework. If running a 4 person mixture, then "low memory mode" can be selected. If needing to ignore loci for analysis, that can be done here.



32.5.4.1 <u>Ignoring Loci</u> (with TL approval)

32.5.4.1.1 Click on "Ignore Loci" in "STR-mix-run settings" window. The window to the left will appear.

32.5.4.1.2 Click on locus you wish to ignore, then click on "arrow right" button to move it to the "Exclude" column. Click "Save" will change loci for this run only. Default settings will not be changed.

32.5.5 After everything has been properly added to the "Configure Analysis" window, click "Confirm".

32.5.5.1 Drag text file with evidentiary file results (the text file exported from Genemarker as in DNA WI-34 step 14) to be deconvoluted

into "Step 2: Add Evidence Profile Data". Conversely, you can click "Add Profile", then "Select Text File" and browse to this file. This file HAS stutter peaks. (The stutter peaks HAVE NOT been filtered out). Browse through the list of samples in that file, click on the one you are deconvoluting, and click "Add".

Confirm that correct sample was chosen, then click on "Add Profile Data".

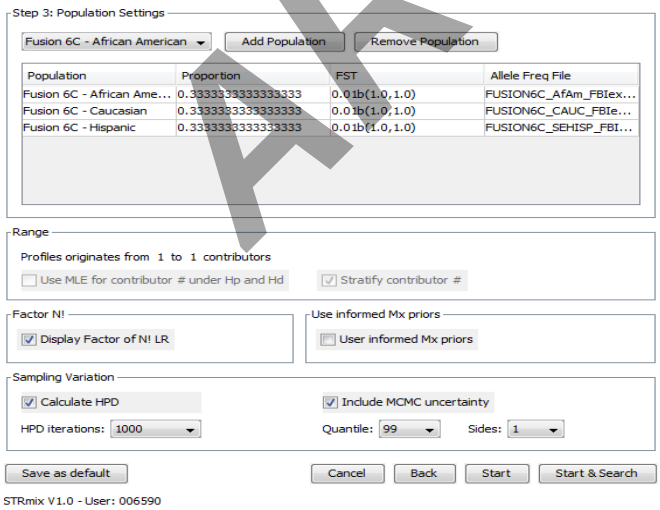
32.5.5.2 When adding a reference profile, repeat the step above, but with the text file containing the known you are adding to a hypothesis.

32.5.5.3 If you are conditioning (a) known(s) to H_2, that profile must be added first. Click on the known in the "Add Reference Profile Data" section then click on "Change H_d" This will add an "x" to the "contributor to H_d" column. After profiles have been conditioned to H_2, knowns only appearing in H_1 can be added. After all knowns have been added and proper hypotheses have been set, click "confirm settings".

32.5.6 When just performing a deconvolution (i.e. not obtaining a likelihood ratio), whether or not a known is conditioned, there is nothing to do in the "Population Settings" window. Click on "Start" to deconvolute the profile.

32.5.6.1 When knowns are applied to any hypothesis, 3 populations must be added to "Step 3: Population Settings": Fusion 6C – African American, Fusion 6C – Caucasian, and Fusion 6C – Hispanic. This setting is completed the first time STRmix™ is run, and need not be completed on subsequent deconvolutions.

32.5.6.1.2 Select each population from the pull-down menu and click "Add population" to populate the window below after each population is selected. Click "Save as default" after all 3 are added so that this step does not have to be repeated with subsequent deconvolutions with the same kit. However, since this is kit specific, it must be repeated the first time a kit is run on STRmix™. The diagram to the left is how the population settings should look, when knowns are applied to any hypothesis, prior to hitting "Start".



32.5.7 When the "Calculation Results" window opens, click on "Run Report", then "create report" in the "Advanced Report v3.0.7" window that then pops up. Save the report

as prompted to the file location defaulted in STRmix™.

32.5.8 Print pages 1 and 3 of the Advanced Report for the case jacket.

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 "LR from Previous Analysis", and navigate to the results file folder of the deconvolution that was previously completed in the F:\Results folder. Click on the "settings" file from that deconvolution.

32.6.2 In the "Configure Analysis" window, the case number shall remain the same. The sample ID will be changed to reflect the calculation currently being completed, (i.e. 1G1 LR 2 changes to 1G1 LR 3). This can be further described in the case notes. Since this is to a previous deconvolution, Step 1, the MCMC settings, are unable to be changed. Click "Confirm".

32.6.3 You will be unable to add or remove evidence profiles. Add reference profiles to the bottom half of the "add profile data" window, as in 32.5.5.2. The known will only be added to H_p. Knowns cannot be added to 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 If this is your first time doing a LR in STRmix™, "Step 3: population settings" will not have any populations. Add populations and confirm settings as in 32.5.6.1. Click "start".

32.6.5 When the "Calculation Results" window opens, click on "Run Report", then "create report" in the "Advanced Report v3.0.7" window that then pops up. Save the report to the file location defaulted in STRmix™.

32.6.6 Print pages 1 and 3 of the Advanced 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 "Add to Batch", and add a sample as you would in 32.5. 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".

32.7.4 If all samples have been added, but you do not wish to start the batch at this time, click "Exit Batch Mode". 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 Batch".

32.7.6 If you wish to end the batch, click "End Batch". The sample that is currently being deconvoluted will run to completion, and then the batch will 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 "Samples In Batch" window, then click "Delete Analysis".

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 Run Information of the Advanced Report

32.8.1 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 acceptable ranges for the quality parameters based on the internal validation are documented on this QR.

32.8.2 **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.

32.8.3 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 1000 warrants additional scrutiny.

32.8.4 Average (log)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.

32.8.5 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, then there exists the possibility that the analysis hasn't converged. If the GR value is above 1.2, the results of the analysis should be scrutinised. Running the analysis for a larger number of iterations may reduce the GR in these instances to below 1.2.

32.8.6 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.

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.

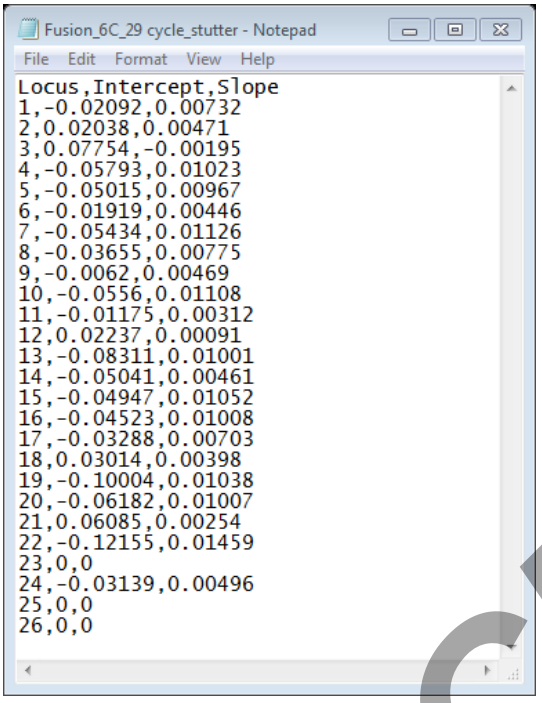
Used in conjunction with the average (log_{10})likelihood, a large allele variance or stutter variance may indicate that the PCR did not perform as expected.

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

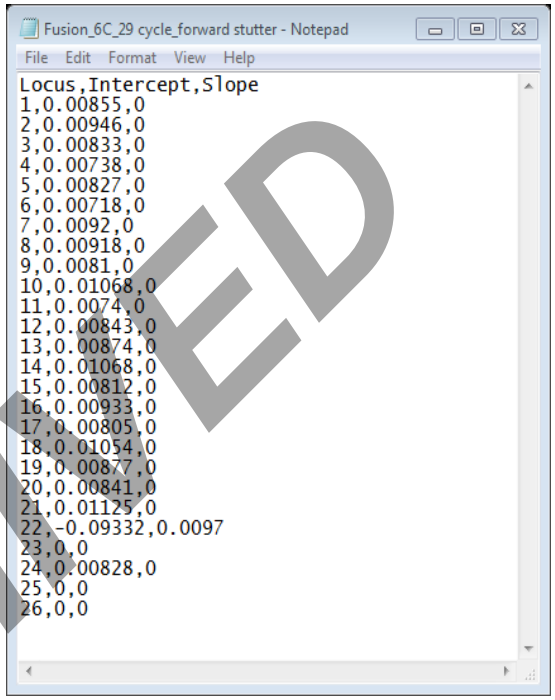
If some data has been omitted or mistakenly left in the STRmix™ input file, this often results in a low average (log_{10})likelihood and high variances.

Appendices

Appendix 1: Stutter File



Appendix 2: Forward Stutter File



Appendix 3: 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	D1S1656	D2S441	D2S1338	CSF1PO	Penta D	TH01	D21S11	D8S1179	D19S433	SE33	FGA
6	0	0	0	0	0	0	0.02126	0	0	0	0	0
7	0	0	0	0	0.02889	0	0.03261	0	0	0	0	0
8	0	0	0	0	0.0446	0	0.03284	0	0	0	0	0
9	0	0	0	0	0.03888	0.02391	0.03904	0	0.05481	0	0	0
9.3	0	0	0	0	0	0	0.01725	0	0	0	0	0
10	0	0	0.04354	0	0.05278	0.02139	0	0	0.06995	0	0	0
11	0	0.06336	0.06581	0	0.06379	0.02226	0	0	0.08066	0.04463	0	0
11.3	0	0	0.0307	0	0	0	0	0	0	0	0	0
12	0	0.0734	0.07349	0	0.08068	0.023	0	0	0.08394	0.0616	0	0
13	0.06355	0.08337	0	0	0.09312	0.02947	0	0	0.07828	0.06884	0.07679	0
13.2	0	0	0	0	0	0	0	0	0	0.07345	0	0
13.3	0	0	0	0	0	0	0	0	0	0	0	0
14	0.07844	0.09366	0.04604	0	0	0.03312	0	0	0.08691	0.07851	0.08063	0
14.2	0	0	0	0	0	0	0	0	0	0.09034	0	0
15	0.09291	0.10519	0.05485	0	0	0.03318	0	0	0.0869	0.08811	0.08763	0
15.2	0	0	0	0	0	0	0	0	0	0.09411	0	0
15.3	0	0.06924	0	0	0	0	0	0	0	0	0	0
16	0.0933	0.11508	0	0.07082	0	0	0	0	0.09698	0.09051	0.10257	0
16.2	0	0	0	0	0	0	0	0	0	0.10265	0	0
16.3	0	0.07731	0	0	0	0	0	0	0	0	0	0
17	0.10154	0.12828	0	0.07211	0	0	0	0	0	0	0.10101	0
17.3	0	0.08733	0	0	0	0	0	0	0	0	0	0
18	0.11101	0	0	0.07885	0	0	0	0	0	0	0.11568	0
18.3	0	0.0956	0	0	0	0	0	0	0	0	0	0
19	0.1177	0	0	0.08786	0	0	0	0	0	0	0.12278	0.05522

19.3	0	0.11022	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0.09115	0	0	0	0	0	0	0.12455	0.0641
21	0	0	0	0.0908	0	0	0	0	0	0	0.1356	0.0703
21.2	0	0	0	0	0	0	0	0	0	0	0.08847	0
22	0	0	0	0.08697	0	0	0	0	0	0	0.13586	0.08002
Repeat	D3S1358	D1S1656	D2S441	D2S1338	CSF1PO	Penta D	TH01	D21S11	D8S1179	D19S433	SE33	FGA
22.2	0	0	0	0	0	0	0	0	0	0	0.09782	0
23	0	0	0	0.09679	0	0	0	0	0	0	0	0.08644
23.2	0	0	0	0	0	0	0	0	0	0	0.10263	0
24	0	0	0	0.10597	0	0	0	0	0	0	0	0.09284
24.2	0	0	0	0	0	0	0	0	0	0	0.1108	0
25	0	0	0	0.12193	0	0	0	0	0	0	0	0.09859
25.2	0	0	0	0	0	0	0	0	0	0	0.10887	0
26	0	0	0	0.11848	0	0	0	0	0	0	0	0.10413
26.2	0	0	0	0	0	0	0	0	0	0	0.12206	0
27	0	0	0	0	0	0	0	0.06409	0	0	0	0.09188
27.2	0	0	0	0	0	0	0	0	0	0	0.12681	0
28	0	0	0	0	0	0	0	0.077798	0	0	0	0
28.2	0	0	0	0	0	0	0	0	0	0	0.13212	0
29	0	0	0	0	0	0	0	0.083498	0	0	0	0
29.2	0	0	0	0	0	0	0	0	0	0	0.13372	0
30	0	0	0	0	0	0	0	0.094359	0	0	0	0
30.2	0	0	0	0	0	0	0	0.082745	0	0	0.14036	0
31	0	0	0	0	0	0	0	0.10219	0	0	0	0
31.2	0	0	0	0	0	0	0	0.086892	0	0	0.15647	0
32	0	0	0	0	0	0	0	0.110109	0	0	0	0

32.2	0	0	0	0	0	0	0	0.096907	0	0	0.15267	0
33.2	0	0	0	0	0	0	0	0.10599	0	0	0	0
35	0	0	0	0	0	0	0	0.088913	0	0	0	0

This page intentionally left blank.