STRmix™ v2.8.0 Release and Testing Report

29 September 2020

Purpose of the document:

This document explains the enhancements to STRmix™ within v2.8.0 and the various activities performed as part of the developmental validation. It also provides users with in-house internal validation and installation guidance.

Description of changes:

A summary of the changes made within STRmix™ v2.8.0 is provided below. Some differences to the *LR* are expected between versions v2.7.0 and v2.8.0 when carrying out an *LR* from Previous (see Science change 1 below).

General use:

1. Improvements to both low and normal memory modes to allow for processing of larger problems with lower RAM requirements
2. Code refinements
3. Improvements to the packaging of the STRmix™ software
4. Improvements to logging
5. Improvements to when warning messages regarding ignoring peaks below AT are shown
6. Continued improvements to error message handling of null pointer and uncaught errors
7. Fix to a compatibility issue with Non-English (French) date format preventing a v2.5.11 *LR* from previous running in the v2.6 series and v2.7.0
8. Fix to compatibility issue with accents (eg é) used in comments
9. Addition of VeriFiler™ Plus kit type
10. Various improvements to aid in developmental testing and validation
11. Improvements in internationalisation to handle grammatical gender usage
12. Change to Model Maker to check for partial references against the kit itself rather than against the evidence input samples
13. Improvements to tickbox setting display in the User Interface
14. Improvement to allow loci to be ignored in *LR* from Previous
15. Allow batch mode to exit on finish (for export from FaSTR™ DNA)
16. Changes to the way retrospective drop-in alleles are calculated for investigations on runs in versions prior to STRmix™ v2.5.11 where this information was not written to the results folder
17. ‘Drop-in frequency’ in user interface changed to ‘Drop-in rate parameter’
18. TM added to STRmix files on disc
19. Updates to website links in About screen, installer and Purchase button
20. Improvements to Auto Database search so that it is prohibited if the setup is fully conditioned
21. Update to the installer background image website
22. Fix to display milliseconds correctly in "human readable" time formatting when execution time is over a day

23. Change to allow Allele frequency files and stutter exceptions files not to contain non-relevant (gender/quality/Y/ignored) loci
24. Updates to the Software Licence Agreements

Science changes:

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

New Features:

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

Changes to reports:

1. Fix to resolve an issue where not all STRmix™ v2.6 series reports were able to be compiled in later versions of STRmix™
2. Internationalisation of reports
3. Removal of Calculation Array Check from extended outputs
4. Improvements to Model Maker extended output file including adding profile indices and filenames
5. Removal of run information from Results.sha512 hash to allow the use of the Results.sha512 as an indicator of same results between two runs
6. Addition of a new audit hash (Samples.sha512)to include input files only

7. Addition of the Minimum Resampled Count per population to PDF report
8. Inclusion of count of log(*LR*)=0 on the H_d True Tester plots in the report
9. Addition of assumed partial loci to the PDF report when conditioning on a partial reference
10. Improvement to *LR* from Previous report: When running an *LR* from Previous, if the seed from the interpretation cannot be found, rather than listing the *LR* as 0, this line is no longer populated in the PDF report
11. Addition of inter-section spacing before the last report component in PDF reports
12. Replacement of 3.2 with 3.3 in CODIS cmf report
13. Inclusion of genotype sets and re-calculated profile probability from burn-in to the Post Burn-in extended output file
14. Improvement to allow the *LR* sorting in the PDF report for Familial searches to handle a blank value
15. Normalisation factor included in *LR* extended output
16. SVG files in the reports replaced with jpeg/png files

Science changes:

A change to allele frequency sampling to use *k*+1 instead of *k* within the posterior mean allele frequency calculation has resulted in changes to the point estimate *LR*s between STRmix versions 2.7.0 and 2.8.0. This change has been shown to be within 1 order of magnitude using deconvolutions carried out in v2.7.0 run in Database Search in v2.7.0 and in v2.8.0 (see Figure 1). The profiles used were generated using GlobalFiler from the publicly available DNA dataset from the PROVEDIt[1] Initiative.

These same samples were also fully deconvoluted and run in Database Search in STRmix v2.8.0 and plotted against the same samples run in v2.7.0 to show the magnitude of changes within STRmix v2.8.0 that have affected both the MCMC and the *LR* (see Figure 2). The differences were as expected.

[1] https://lftdi.camden.rutgers.edu/provedit/files/

Plot of *LR* differences:

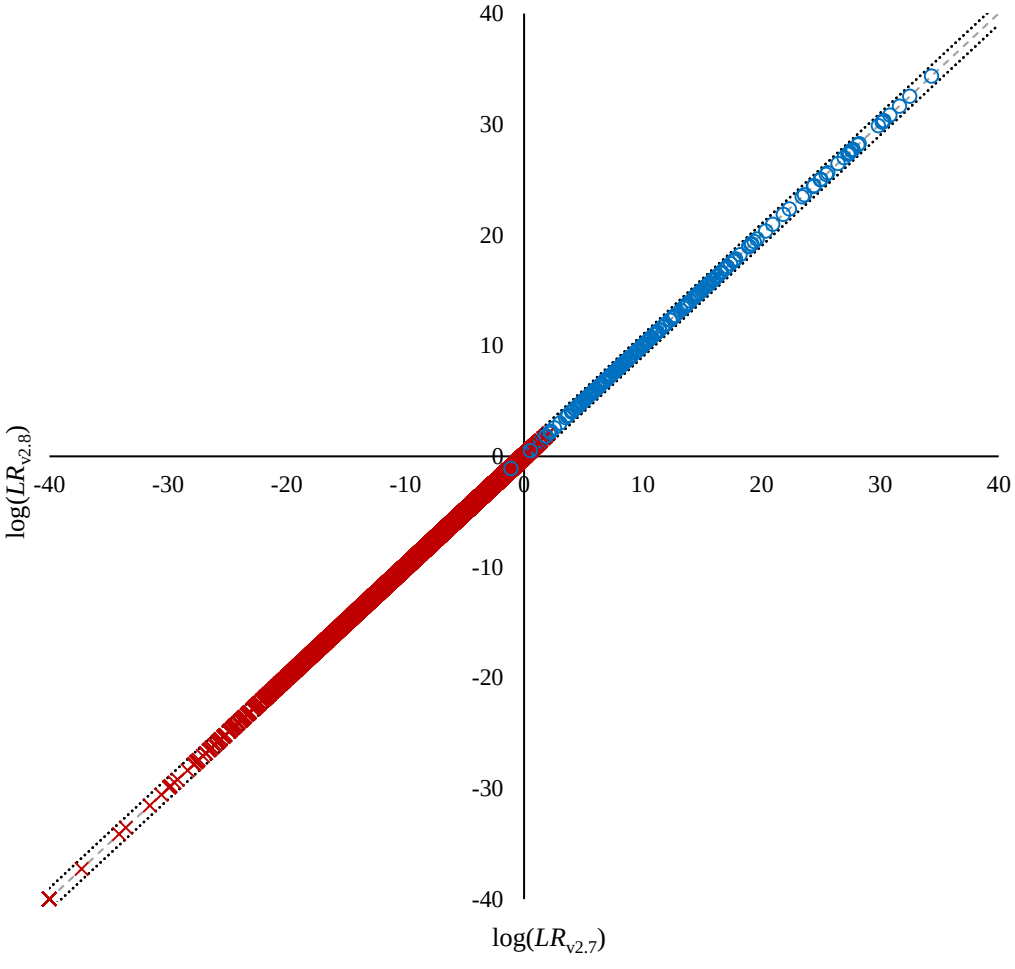

Figure 1: Plot of log(*LR*) from v2.7.0 vs log(*LR*) from v2.8.0 (using *LR* from previous) for single source and mixed DNA profiles that were interpreted in STRmix v2.7.0.

Plot of overall differences:

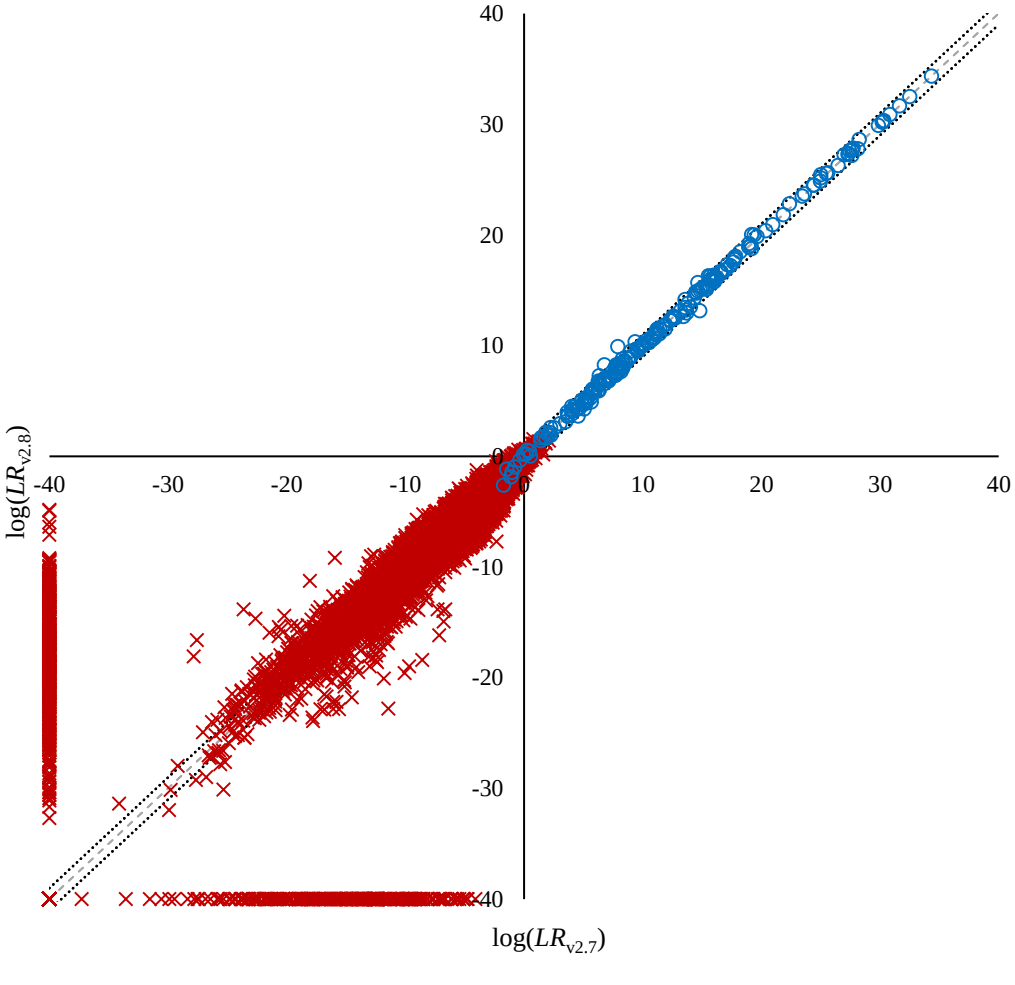

Figure 2: Plot of log(*LR*) from v2.7.0 vs log(*LR*) from v2.8.0 when samples have been deconvoluted and run in Database Search in both versions of STRmix™.

Testing scope:

In scope: Functional testing for the following modules was undertaken:

- Security (Administration mode)
- Interpretation (profile deconvolution)
- Variable Number of Contributors (varNOC)
- Multi kit inputs and stitched reference profiles
- Likelihood ratio calculation including unrelated, relatives, sub-source and sub-sub-source *LR*s, unified, stratified, HPD, and under variable number of contributors calculations
- Database and Familial Search functionality and *LR*s
- Model Maker
- Batch Mode

- *LR* from Previous
- Investigation Batch (*LR* from Previous and Database Search)
- Hd True Tester
- Top Down Approach
- Reporting module including CODIS xml and summary reports
- Biological models using extended output reports

Out of scope: Nothing

Summary of testing:

Security/Administration

- The ability to set a password (or not use a password) was tested along with restriction of access to STRmix™ settings. All areas functioned as expected.

Multiple chains and threads and setting the seed

- Runs were tested using 4, 8 and 20 chains. Results achieved were similar however run time increased as expected with more chains.
- When more threads were run (within the capacity of the computer), run time was reduced and more memory consumed as expected.

Extended output calculations

- Extended outputs were examined from a number of loci/profiles covering a range of scenarios. Calculations were replicated in Excel 'by hand' including total allelic product, expected allele and stutter peak heights, smart start, log(likelihood) values, shifted log normal penalties, replicate and kit efficiencies, drop-in and other penalties. All calculations were found to be as expected

- Run diagnostics, weights, and step sizes were calculated in Excel 'by hand' from the extended outputs with results as expected.

Variable numbers of contributors

- VarNOC calculations (including multi-kit scenarios) were replicated in Excel 'by hand' from the extended outputs with results as expected.

Likelihood ratios

- Various scenarios were tested (number of contributors, with and without assumed contributors) and all *LR* results (locus-specific point estimates and relatives point estimates) were as expected.
- Regression testing included checking *LR*s for both unrelated and related hypotheses (siblings, parent/child, cousins, half sibs, uncle/nephew, grandparent/grandchild) against v2.7.0.
- 'By hand' calculation of point estimate, HPD, stratified, and unified (including stratified unified HPD) *LR*s for interpretations assigning both a single number of contributors and a range of contributors (varNOC).

Model Maker results and diagnostics

- Outputs were examined from the Model Maker extended outputs. All formulae were calculated in Excel 'by hand' from the extended outputs with results as expected.

Database Search

- 'By hand' testing of familial search *LR*s for interpretations assigning a range of contributors (varNOC). Results obtained were concordant. Database search *LR*s were compared against 'by hand' tested *LR*s and results were concordant.
- Functionality testing included changing database, F_{ST} and sub-source *LR* calculations, allele frequencies, familial mutation rate and default save settings.

Investigation Batch

- Tested ability to carry out one interpretation against a single or multiple reference/s and also multiple interpretations against multiple references. All possible combinations were calculated. Results obtained were as expected. Likelihood ratios including HPD obtained were concordant within v2.8.0.
- Tested ability to carry out one or more interpretations against one or more databases as batches. Results obtained were as expected. Likelihood ratios were concordant within v2.8.0.

H_d True Tester

- H_d True Tester calculations for a one locus single source profile with ambiguity and one locus two person mixed DNA profile were replicated 'by hand'. The results and plots were as expected.

Top Down Approach

- The Top Down Approach was tested on mixed DNA profiles ranging from 2 person to 5 person. Results were obtained using three different DNA profiling kits with various settings for Database Search used. In each instance step sizes were correctly calculated and the results for each AT were the same as those obtained when the sample was run with the same AT as an individual deconvolution.
- *LR*s obtained from the Top Down Approach were verified to be the maximum obtained from all the Database Searches in the steps as expected.

Run conditions

- Functionality testing included testing a range of different profiles (including different numbers of contributors up to five person mixtures and different template amounts) and reviewing *LR*s calculated for both H_p and H_d true scenarios.
- Further functionality testing included checking *LR* from Previous, Database Search, Mx priors, HPD, and Batch Mode.

All tests gave the expected results.

Compatibility with previous versions:

Compatibility testing was undertaken on kits (including stutter files), allele frequency files and databases transferred from STRmix™ v2.5.11 to v2.8.0. In addition, *LR* calculations were undertaken in v2.8.0 on deconvolutions carried out in v2.7.0.

Kits created in STRmix™ v2.5.11, v2.6.0 to v2.6.3 and v2.7.0 are compatible with v2.8.0. However, if using a v2.5.11 kit, note that forward stutter variance is no longer bound to allele variance and has its own field in the kit set up.

LR from Previous and Database Search analyses were carried out using deconvolutions undertaken in STRmix™ v2.3.10, v2.4.08, v2.5.11, v2.6.0 and v2.7.0. The results obtained were as expected. *LR* from Previous and Database Search analyses cannot be carried out on deconvolutions undertaken in versions prior to v2.3.06.

If calculating an *LR* from Previous within STRmix™ v2.8.0 using a deconvolution from a previous version of STRmix™, population and kit names must be the same between versions.

Instructions for users:

1. Always ensure you install the latest version of STRmix™ as soon as practicable.
2. Ensure that your installation is on the same PCs currently running STRmix™ (existing users).
3. You will need to obtain a new licence for v2.8.0. Do this by copying the code that appears on starting STRmix™ v2.8.0 and email to support@strmix.com.
4. Follow the steps in the *STRmix™ v2.8 Installation Manual* - please email support@strmix.com if any technical issues are encountered.
5. Kit and stutter files are concordant between STRmix™ v2.7.0 and v2.8.0 with the following exceptions:
6. If upgrading from STRmix™ v2.7.0, then Model Maker will not need to be re-run in STRmix™ v2.8.0.
7. When re-running Model Maker ensure drop-in modelling has been disabled. Please refer to the *STRmix™ v2.8 Implementation and Validation Guide* for details.
8. If you are using STRmix™ for the first time, please refer to the most recent *STRmix™ v2.8 Implementation and Validation Guide* for validation guidance.
9. For upgrades to v2.8.0 we recommend that you undertake an in-house validation prior to use of STRmix™ v2.8.0 in casework. A suggested performance check for an upgrade from v2.7.0 to v2.8.0 involves the interpretation of fifty profiles of varying quality (template) and varying numbers of contributors using new stutter files (if appropriate) and Model Maker results. A suggested plan is:
   a) An unambiguous (high template) single source profile where weights = 1 for a single genotype will result in point *LR*s within one order of magnitude difference (including relative propositions) using the same allele frequency database and theta values due to allele frequency sampling to use *k*+1 instead of *k* within the posterior mean allele frequency calculation.
   b) Mixed DNA profiles that contain multiple low level (non-assumed) contributors, where multiple genotype sets with dropped alleles for both contributors are being considered. This will result in different but similar *LR*s due to the expected variability within the MCMC, modelling changes (*k*+1), and changes to the calculations in v2.8.0.
   c) Mixed DNA profiles where one contributor is a trace or minor contributor with alleles in stutter positions (back, forward, double back etc.) of the major contributor should be interpreted and the results be intuitive.
   d) *LR* from Previous interpretations on deconvolutions carried out in previous versions should result in the same *LR* within one order of magnitude difference due to allele frequency sampling to use *k*+1 instead of *k* within the posterior mean allele frequency calculation.

   Due to science changes differences between weights, *LR*s and HPD calculations are to be expected.